Cognitive Psychology Unit

Consolidation in Visual Working Memory: How Does it Operate and Can it Be Facilitated?

Thesis (Cumulative Thesis)

Presented to the Faculty of Arts and Social Sciences

of the University of Zurich

for the degree of Doctor of Philosophy (PhD)

by Clara Overkott

Accepted in the fall semester 2020

On the recommendation of the doctoral committee composed of

Prof. Dr. Klaus Oberauer (Main Supervisor)

Prof. Dr. Mark Nieuwenstein

Zurich, 2020

Abstract

Consolidation is an attentionally-demanding process and requires time to create a stable representation in in visual working memory (VWM). The first goal of my dissertation was to investigate how consolidation in VWM operates and whether it can be facilitated. Therefore, my first goal was to test whether consolidation operates as a ballistic process or not. The data support the notion that consolidation is not a ballistic process, rather, consolidation is a graded process that is under strategic control and can be interrupted. Consolidation time was beneficial under verbal suppression, ruling out the possibility that the free time given for consolidation was used for labeling. My second goal was to test whether labeling could be used to facilitate consolidation. Additionally, I tested whether a potential labeling benefit in VWM would translate into visual long-term memory (VLTM) in study 2. The results revealed that labeling is beneficial for the retention of fine-grained information in VWM, and only limitedly benefits categorical VLTM. Lastly, I tested whether labeling can lead to a cost, by filtering the non-labeled item feature in study 3. The results show that labeling produced asymmetric effects, as labeling always benefitted the detailed information of these features, but this came at the cost of some of the detailed features of the non-labeled items. I conclude that labeling can be used to facilitate consolidation, but the benefit is subject to limitations.

Zusammenfassung

Die Konsolidierung ist ein aufmerksamkeitsintensiver Prozess und benötigt Zeit, um eine stabile Gedächtnisrepräsentation im visuellen Arbeitsgedächtnis (VAG) zu bilden. Das erste Ziel meiner Dissertation war es zu untersuchen, wie die Konsolidierung im VAG funktioniert und ob sie erleichtert werden kann. Daher bestand mein erstes Ziel darin zu testen, ob Konsolidierung als ein ballistischer Prozess funktioniert oder nicht. Die Daten unterstützen die Vermutung, dass Konsolidierung kein ballistischer Prozess ist, sondern dass Konsolidierung ein abgestufter Prozess ist, der unter strategischer Kontrolle steht und unterbrochen werden kann. Die Konsolidierungszeit war auch bei verbaler Suppression vorteilhaft, so dass die Möglichkeit ausgeschlossen werden konnte, dass die für die Konsolidierung gegebene freie Zeit für Verbalisierungen genutzt wurde. Mein zweites Ziel war es, zu testen, ob die Verbalisierung zur Erleichterung der Konsolidierung genutzt werden kann. Zusätzlich testete ich in Studie 2, ob sich ein potenzieller Nutzen der Verbalisierung im VAG ins visuelle Langzeitgedächtnis (VLZG) transferieren liesse. Die Ergebnisse zeigten, dass die Verbalisierung für die Beibehaltung detaillierter Informationen im VAG vorteilhaft ist, aber nur begrenzt zu mehr kategorischer Informationen im VLZG führt. Schliesslich testete ich in Studie 3, ob die Verbalisierung zu Kosten führen kann, indem die nicht-verbalisierte Eigenschaft eines Objekts gefiltert wird. Die Ergebnisse zeigen, dass die Verbalisierung asymmetrische Effekte erzeugte. Die Verbalisierung kam immer den detaillierten Informationen der Merkmale zugute, führte jedoch zu Kosten der detaillierten Merkmale der nicht-verbalisierter Eigenschaften für einige Objekte. Ich komme zu dem Schluss, dass die Verbalisierung zur Erleichterung der Konsolidierung eingesetzt werden kann, dieser Nutzen aber Einschränkungen unterliegt.

Contents

PART I - SYNOPSIS

1. Introduction

Every moment we are exposed to numerous visual information of our surroundings. The cognitive system responsible for holding and maintaining such visual information is called *visual working memory* (VWM) (Nelson Cowan, 1988; Oberauer, 2009; Oberauer & Hein, 2012). VWM has a limited capacity and can only maintain a small amount of information simultaneously for ongoing processing (Luck & Vogel, 2013; Oberauer et al., 2016). This is in contrast to visual long-term memory (VLTM), which stores an unlimited amount of visual information over long periods for up to months and years (Brady et al., 2008; Konkle et al., 2010a, 2010b). Previous research has shown that a free time interval provided immediately after the presentation of the memoranda can be helpful to create more robust representations, thereby overcoming the capacity limitations in VWM (Ricker & Cowan, 2014). This benefit was associated with consolidation, but the mechanisms of this process yet remain unclear. Likewise, it has been hypothesized that this free time is used for labeling, thereby possibly benefitting VWM and VLTM. For example, recent research showed that verbally labeling the memoranda benefits VWM (Forsberg et al., 2019, 2020; Souza et al., 2020; Souza & Skóra, 2017).

To better understand the mechanisms of the consolidation process, my first research question of this thesis is as follows: How does consolidation in VWM operate? To answer this question, I will first explain in the following sections how representations are built in VWM and discuss capacity limitations in VWM in relation to time. I will summarize why (A) consolidation time is necessary to build a representation and how it may operate when time is limited. I will lay out that (B) verbal labeling manipulated during consolidation benefits memory. This raises the possibility that time given to consolidate is used exclusively for labeling, suggesting that consolidation and labeling are the same. In answering my first research question in study 1, I could show that time to consolidate is beneficial even when labeling is inhibited. This leads me to my second research question of this thesis: Can verbal labeling be used to facilitate consolidation? Furthermore, can this translate into a VLTM benefit (study 2) and will it come at a cost (study 3)? Finally, I will provide more detailed information about the three studies and summarize their findings.

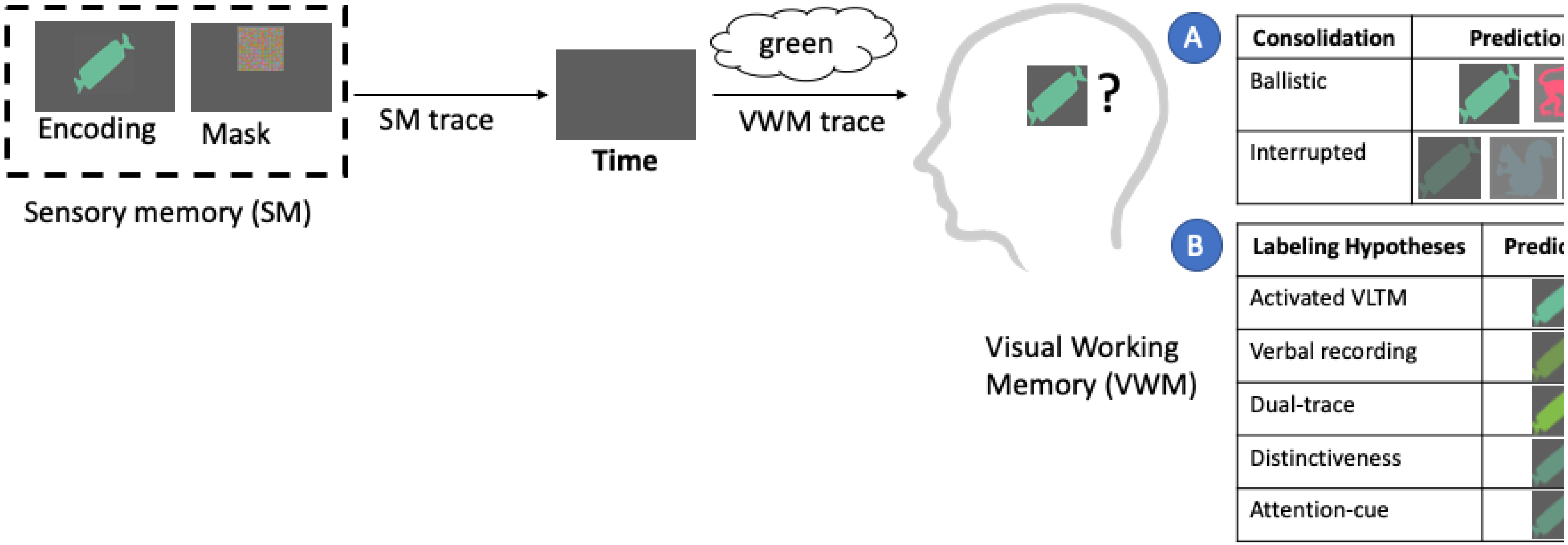
1.1. How Is a Memory Representation Created?

To investigate the benefits of time during the formation of a memory representation, it is important to understand which processes are involved in the creation of a memory representation. Figure 1 schematically shows the process of creating a representation from the first encounter in sensory memory to a stable representation in VWM. When the stimulus of a mint-green candy is presented, it is encoded as a sensory trace during a small amount of time (Turvey, 1973; Vogel et al., 2006). This visual trace is highly susceptible to interference of any other incoming sensory information (Massaro, 1975; Sperling, 1960). Hence, this sensory memory trace needs to be translated into a stable VWM representation – a process which requires attention and time and is achieved through consolidation (Ricker, 2015).

Encoding and consolidation both have a distinct function in building a memory representation. Still, these two processes have often been confused or used interchangeably (Vogel et al., 2006; Woodman & Vogel, 2005, 2008), which may be one of the reasons why consolidation has only limitedly been investigated so far. However, today there are means to tease these processes apart: Encoding is interrupted by the presentation of a mask, as it disrupts the sensory trace and only leaves consolidation to occur (Ricker, 2015; Ricker & Sandry, 2018). One limitation of this method is the possibility that consolidation already starts during the encoding process. Yet, to date the mask provides the best way of investigating the consolidation in isolation of encoding.

Figure 1

Flow of Creating a Memory Representation Along with the Predictions of How Time Can be Beneficially used for Consolidation or Labeling



1.2. Limitations in VWM

VWM is strongly limited in the amount of information that can be hold and maintained (Nelson Cowan, 2010; Luck & Vogel, 2013; Oberauer et al., 2016). Limitations on how much information can be retained in VWM have been observed in terms of how much information can be stored (quantity) as well as the precision (quality) of information (Zhang & Luck, 2008). In particular, this was used in continuous reproduction tasks, where participants are asked to recall a feature (e.g. color) with a continuous scale (Prinzmetal et al., 1998a; Wilken & Ma, 2004a; Zhang & Luck, 2008). For example, participants will be asked to remember the color of a set of objects, such as a green candy or a pink monkey (for example, Brady et al., 2013). At test, the object is presented in grey surrounded by a continuous color wheel. Participants are instructed to select the precise color of the object by moving the mouse along the wheel. Task performance can be modeled as a mixture of responses around the correct feature space as well as its precision, which is the variability around this value, and guessing. More recently, it has been shown that in remembering these visual features, participants rely on either categorical information (e.g. how close this color is to a canonical “green”, like a green lawn) or continuous information (e.g. a vivid representation of the hue, such as a green-mint color; Bae et al., 2015; Hardman et al., 2017; Pratte et al., 2017). These recent models incorporated the storage of categorical and continuous information as well as its precision into modeling fidelity tasks to explain how VWM representations are limited.

VWM is especially limited in the case when the pace in which visual information is presented to us is very high. For example, when one memory item is presented and followed by a second one in very close time proximity, this second item will be missed, as there is not enough time to reallocate attention to encoding and consolidating the new information (for a review, see Dux & Marois, 2009). This effect has been referred to as attentional blink (Chun & Potter, 1995; Shapiro et al., 1997; Wyble et al., 2009, 2011, 2015). Reason for this is that under these time constraints, attention that can be directed to one item is limited (see also, Oberauer, 2019).

One model attempting to explain the findings of the attentional blink is the episodic simultaneous type-serial token (eSTST) model (Wyble et al., 2009, 2011; see also, Bowman & Wyble, 2007). This model incorporates encoding of information, its attentional selection and consolidation. The model assumes that incoming information activates a representation referred to as type. This type is then bound to a token when its activation level reaches a certain activation threshold. The token is a VWM representation that sustains the activation over time so that it can be retrieved later. This binding process is consolidation. The model further assumes that relevant information to a task is activated by a transient attentional boost, increasing its chances of being bound to a token and thus consolidated. At the same time, the type-token binding (consolidation) suppresses the activation of new incoming information to sustain its own consolidation. In this case, attention cannot be reallocated towards a new incoming type, yielding an attentional blink.

1.3. (A) Time Used for Consolidation in VWM

Previous research has shown, that in case more time is provided after the presentation of a stimulus, memory is better (Ricker & Cowan, 2014). This was attributed to consolidation, which benefits memory as long as attention dwells on the memory item and enough time is provided to finish this process (Bayliss et al., 2015; Kandemir et al., 2017; Nieuwenstein & Wyble, 2014; Ricker & Cowan, 2014; Ricker & Hardman, 2017). It has been shown that time to consolidate depends on the stimulus materials and task set-ups, varying between 500 and 1000 ms (see, Ricker et al., 2018). Ricker and Hardman (2017) showed that more consolidation time was beneficial for the retention of both categorical and continuous memory in a delayed-estimation task. If not enough time for consolidation is provided, memory performance on a subsequent test suffers, compared to when enough time is provided.

There are currently two views in explaining the time constraints of consolidation and they differ in regard to how attention is allocated to proceed consolidation: One view states that attention remains on the item that is currently being consolidated, even if this comes at the expense of new information. The second view assumes that in case ongoing consolidation is interrupted, attention is reallocated to the new information that is then consolidated.

The first view is based on the *strict attentional bottleneck model* (Jolicœur & Dell'Acqua, 1998), which assumes that attention is a limited, central resource that can only work in a serial manner. Thus, attention can only be directed to new information once consolidation has ended. In this case, consolidation follows a ballistic manner, meaning that once consolidation has started it has to finish (Ricker, 2015; Ricker & Hardman, 2017). This will come at the expense of any incoming information that is presented during the consolidation of an item.

Let us assume the following scenario: Three pictures are presented at a fast pace so that one cannot consolidate them all. The attentional-bottleneck model assumes that the first picture of a "green candy" will be fully consolidated at the expense of the second picture a “blue squirrel”. By the time the third picture of a “pink monkey”, consolidation of the first picture should be done and the third picture can be consolidated at the expense the fourth picture. According to the ballistic nature of consolidation in this model, a fully consolidated memory trace of the green table and pink monkey will be retained in VWM, but lack of a memory trace of the blue squirrel. This is exemplified in Figure 1 (right side under ballistic).

In contrast, in the second view consolidation does not follow a ballistic manner and can be interrupted, which is based on the view of a *resource-sharing model of central attention allocation* (Lehle & Hübner, 2009; Tombu & Jolicœur, 2003). Similarly to the bottleneck model, it assumes that attention is a limited resource, but it can be strategically distributed across multiple incoming information presented in very close time proximity (e.g. simultaneously) and then allows parallel processing. This model predicts that consolidation can be interrupted (Bayliss et al., 2015; Kandemir et al., 2017; Nieuwenstein & Wyble, 2014). In this scenario, the first picture will be consolidated until it is interrupted with the presentation of the second picture. Likewise, the second picture will be consolidated until interrupted by the third and so on. In the end, three picture representations are maintained in memory that have not been fully consolidated, but are less robust (as depicted in more blurred pictures in Figure 1, right side under interrupted).

1.4. (B) Time Used for Verbal Labeling

We also see that when enough time for consolidation is provided, people benefit from verbal labeling in VWM (Forsberg et al., 2019, 2020; Souza et al., 2020; Souza & Skóra, 2017). Studies investigating the role of verbal labeling typically instruct participants to label the item during or immediately after the presentation of the memoranda when consolidation occurs (for example, Lupyan, 2008; Richler et al., 2013; Souza & Skóra, 2017). On the one hand, this could imply that time to consolidation was used for labeling, raising the hypothesis that consolidation and verbal labeling reflect the same process. On the other hand, there is the possibility that labeling may facilitate consolidation in building a stable memory representation.

To better understand the effect of labeling, I will present five current hypotheses on how labeling influences the retention of information in memory (see also study 2 and 3). Importantly, most hypotheses gain support from both memory systems and do not predict that labeling would have different effects on VWM and VLTM. This suggests that for the case that labeling facilitates consolidation, it should benefit consolidation for both memory systems. I will additionally lay out how these hypotheses relate to consolidation.

One hypothesis is the *activation of categorical VLTM* hypothesis assumes that two visual memory traces are produced, one from the sensory item and one from the VLTM representation of the category activated by the label, thereby activating more continuous information (Lupyan, 2012; Souza & Skóra, 2017). For example, Souza and Skóra (2017) sequentially presented four continuously varying colored discs that were all reproduced at the end of a trial with the use of a color wheel. Participants were instructed to either say the color aloud (aka labeling) or repeatedly say “bababa” (aka suppression). Labeling clearly increased performance and this was especially due to more fine-tuned responses (continuous) about the color hue or its precision. This is depicted in Figure 1, where the candy clearly contains a lot of continuous information of the color mint. In contrast to the benefits of consolidation time (more categorical and continuous memory), this hypothesis predicts an increase in continuous memory. This gives a hint that labeling may facilitate consolidation by an additional increase in continuous memory.

In contrast, there are also labeling hypotheses predicting a cost on memory. The *verbal recording* hypothesis assumes that the label activates a verbal trace (Alogna et al., 2014; Donkin et al., 2015; Lupyan, 2008; Schooler & Engstler-Schooler, 1990; Sense et al., 2017) that can come at the expense of the visual trace, resulting in a cost for visual memory (Kelly & Heit, 2017; Souza & Skóra, 2017). Evidence for this hypothesis stems from the verbal overshadowing effect, in which a label activated a category that interfered with recalling the color. In this scenario, the precise information of the color a hue was no longer present in VLTM as it was overshadowed by the category label (Alogna et al., 2014; Brandimonte et al., 1992, 1997; Lupyan, 2008; Schooler & Engstler-Schooler, 1990). Thus, the activation of the verbal category label results in the loss of continuous information, which is depicted in Figure 1. This is in contrast to what is predicted by more consolidation time, which benefits memory. Thus, the label may have resulted in a cost for consolidation of more fine-tuned information. In this case, labeling does not facilitate consolidation, but hampers it.

Labeling was further explained as a *dual trace* hypothesis, wherein a verbal trace of the label and a visual trace of the sensory input information are built, thereby providing additional categorical information to memory (Souza & Skóra, 2017). This is in contrast to the activation of categorical VLTM hypothesis, where two visual traces are built. Evidence stems from studies showing that the verbal overshadowing effect could be reversed in VLTM (Brandimonte et al., 1997; Brown et al., 2014). For example, when participants were asked to give a detailed description of previously learned easy-to label pictures, recall performance in a featural memory test increased. This hypothesis predicts that more categorical knowledge about the color is added to the memory representation, resulting in a very clear prototypical green color, which is shown in Figure 1. This hypothesis suggests that the label further facilitates consolidation by increasing categorical memory.

Another hypothesis is the *distinctiveness* hypothesis, which suggests that unique labels make a representation more distinct (Blanco & Gureckis, 2013; Kelly & Heit, 2017; Richler et al., 2013; Souza et al., 2020; Souza & Skóra, 2017), by providing an additional retrieval cue to augment encoding specificity (Blanco & Gureckis, 2013; Richler et al., 2011; Tulving & Thomson, 1973). In a recent study of our lab, we showed that performance in a delayed estimation task increased monotonically with specific labels for continuously varying colors and shapes, reflected in more continuous precision, whereas broad labels reduced precision (Souza et al., 2020). In Figure 1 we can see that this results in a more precise representation of the candy, reflected in a minty and less greenish color. This suggests that unique labels facilitate consolidation by increasing the precision of the memory representation.

Lastly, labels can serve as a *cue to focus attention* to certain aspects of the visual input. This is however only beneficial if the label guides attention to the relevant features of an item. For example, Kelly and Heit (2017) found that color labeling during study reduced the bias towards the color category for VLTM in contrast to an animacy or preference rating during study, but this benefit vanished when participants were informed about the relevant feature of the VLTM test before study. In Figure 1, we can see that this hypothesis helps building a more minty representation – under the assumption that attention was not yet directed to this picture. In this case, the benefit is similar to the one reported by consolidation. Yet, if the task is not relevant, the label guides attention to the memory object, enhancing its chances of being consolidated.

Most hypotheses gain support from both memory systems and predict similar effects on VWM and VLTM, even though the role of labeling on both memory systems has not been investigated in direct relation. This assumption fits with memory models assuming a tight link between the two systems, where VWM is embedded in VLTM (Cowan, 1988; Oberauer, 2002, 2009; Oberauer & Hein, 2012). In these models, VWM consists of a focus of attention and an activated part of VLTM. In Cowan (1988), the focus of attention was described as holding about a handful of objects, whereas in (Oberauer & Hein, 2012) the focus of attention consists of a broad focus as in Cowan (1988) and a narrow focus of attention holding only a single object, thereby giving it a special role. Moreover, the activation of categorical VLTM hypothesis nicely fits within these models, as the label activates a representation in VLTM, thereby facilitating bindings in the focus of attention. Similarly, consolidation activates a VWM representation in the focus of attention.

Most hypotheses imply that labeling facilitates consolidation in VWM and likewise in VLTM. Recent behavioral data, however, supports the categorical activation hypothesis for VWM (Souza & Skóra, 2017) and the attentional cue hypothesis for VLTM (Kelly & Heit, 2017). The former predicts a beneficial effect of verbal labeling by an increase in the fine-tuned information about the feature space, and the latter an absence of this effect in VLTM when the memory item is relevant to the task. Thus, there are contradictory behavioral results in regard to whether labeling is similarly beneficial for the two memory systems. This is in contrast to the five hypotheses that make no clear distinctions about their effect on VWM and VLTM.

1.5. The Present Thesis

Consolidation is essential during the formation of a memory representation as it translates a sensory memory trace into a stable VWM representation. Yet, it is unclear how consolidation operates. It is however known that labeling during the time a memory item is consolidated benefits memory. This raises two possibilities: either consolidation and labeling reflect the same process or labeling can be used to facilitate consolidation.

How does consolidation operate? This is the first research question of the present thesis, aiming to understand consolidation better. In study 1, I will more strongly test for the nature of consolidation to answer the question whether it is a ballistic process or not. Additionally, I will test for the possibility that consolidation time is used for verbal labeling. In the critical Experiment 1, I found that time given to consolidate the memoranda was helpful under a suppression condition, inhibiting labeling. This supports the notion that consolidation and verbal labeling are not the same process and that consolidation operates independently of labeling.

Once I understood better how consolidation operates, I asked the second question of this dissertation: Can labeling be used to facilitate consolidation? In study 2, I tested whether a labeling benefit in VWM could further translate into a VLTM benefit. If so, this would suggest that labeling may also benefit long-term consolidation. In study 3, I tested whether labeling one item feature in VWM would lead to a cost of the non-labeled item feature. This would suggest that labeling may also induce a cost on consolidation.

2. Summary of Study 1

The goal of study 1 was to see how consolidation operates. We aimed to more directly investigate into the nature of consolidation, by examining whether it is a ballistic process.

The general task set-up of four preregistered experiments consisted of a sequential presentation of four visuo-spatial memory items at different locations of the screen, followed by a short or long consolidation interval and a serial recall test for all memory items of a trial. To answer the question whether consolidation could be interrupted, the following variations in the experimental design were implemented: In Experiment 1, (a) recall order was made unpredictable, additionally implementing a random recall test, where participants could no longer prioritize on the first item as in serial order, as this may have provided an incentive to not interrupt consolidation. In Experiment 2, (b) we used a fixed encoding location at the center of the screen, in order to test whether consolidation could be interrupted when shifts of visuospatial attention were no longer necessary. In Experiment 3, (c) we used a distractor task that followed each item to test whether a distractor interrupted consolidation. In Experiment 4, (d) we presented items simultaneously, combined with a distractor task.

Specifically, when conditions of unpredictable recall order (Experiment 1) as well as when shifts of spatial attention at encoding (Experiment 2) were required, the data pattern reflected an all-or none consolidation process. In contrast, results from Experiment 3 and 4 showed that responding to a distractor task interrupted consolidation (c+d), thereby challenging the view that consolidation is a ballistic process.

These finding support the notion that attention is a limited resource that can be strategically allocated to incoming information. This is a crucial point, as it reveals that the true nature of consolidation is a graded process that may be subject to strategic choices of a person. Consolidation is interrupted if a following item or task requires a certain amount of immediate attention, otherwise the payoff may not be motivating enough and consolidation is not interrupted.

In Experiment 1, a suppression condition was included to rule out that verbal labeling is essentially used during the free time interval provided for consolidation. This was not the case as the consolidation time benefit remained under suppression. Subsequently, we did not control for verbal suppression in the remaining experiments of study 1. Interestingly, the serial position curves in the critical Experiment 1 with suppression showed a more linear performance pattern across the serial input positions, whereas the pattern was more u-shaped or flatter serial position curves in the remaining experiments not inhibiting labeling. This suggests that participants may have used the time to consolidate (2000 ms) to further label the memory item. In this case, labeling can assist consolidation, thereby making the memory representation more stable. I then turned to address whether labeling facilitates consolidation of a memory representation.

3. Summary of Study 2

The goal of this study 2 was to put the scope of verbal labeling into a broader perspective by investigating how verbal descriptions affect visual memory over the short- and long-term.

This question was addressed across four preregistered studies, in which clip-art objects were presented in continuously varying colors. Memory test consisted of color recall that occurred after the presentation of three items (VWM) and at the end of the experiment (VLTM). During item presentation, labeling opportunities were manipulated by instructing participants to say (a) the color, (b) the object, (c) the color-object association aloud, or (d) say “bababa” aloud - which served as a control condition inhibiting verbal labeling.

The results revealed that (a) color labeling (Experiment 1+2) and (c) color-object labeling (Experiment 3) improved VWM performance. In contrast, (b) object labeling (Experiment 1) impaired VWM performance. This latter finding provides a first hint that labeling is not necessarily related to a beneficial effect in VWM and can under some circumstances result in a cost for VWM.

In contrast to VWM, VLTM was not affected by (a) color labeling benefit nor (b) object labeling. In Experiment 2, each VWM trial was successively repeated over three trials to rule out that a lack of a labeling benefit was due to poor VLTM performance. This promoted VWM and VLTM learning, but color labeling remained unchanged for VLTM. Additionally, labeling did not change the learning rate over the three repetitions. These findings support the assumption that verbal labeling operates over short time scales only. Yet, there was a VLTM labeling benefit when (c) the color-object association was labeled (Experiment 3).

Moreover, mixture modeling revealed that beneficial effects in VWM were particularly due to an increase of more detailed representations and, if a long-term benefit was found, it was due to an increase in categorical memory. This dissociation between the two memory systems may be due to the fact that these detailed representations are not retained over time or they are susceptible of interference that accumulated in VLTM.

4. Summary of Study 3

Experiment 1 of study 2 revealed that labeling can also be associated with a cost. Labeling the object of a colored clip-art object induced performance cost in recalling the color from VWM. This cost is predicted by the attentional-cue hypothesis, as verbal labeling directed attention to the object, thereby reducing attention that could be directed to the color of the item. Mixture modeling revealed that this was due to both a reduction in categorical and continuous memory. Thus, information about the color feature was involuntarily filtered with the labeling of the object's identity. The goal of study 3 was to test more directly whether the labeling of one feature would also yield to the involuntary filtering of the non-labeled feature in VWM. If this were the case, it would suggests that time provided for verbal labeling cannot simply be associated with a benefit.

To test this, we presented multi-feature items, varying in color, orientation and spatial frequency. Critically, verbal labeling was manipulated by instructing participants to label the (a) color, (b) orientation or (c) frequency, compared to (d) a baseline suppression condition inhibiting verbal labeling. In Experiment 1, colored triangles were presented varying in their color and orientation. We found that labeling resulted in an increase in the continuous feature information. We further found that labeling the color of a triangle did not result in filtering of the orientation feature (there was some evidence for a cost in recall error in Experiment 1b), whereas orientation labeling led to filtering of continuous color. In Experiment 2, we added an additional spatial frequency feature by presenting Gabor patches to test whether the color feature is special in the way that it is always filtered with labeling of another feature or whether orientation feature is special and is not filtered with the labeling of another feature. The results of Experiment 2 show that color labeling did not lead to the filtering of the frequency information, but frequency labeling resulted in the filtering of continuous color and orientation information.

Hence, labeling seems to have asymmetric effects: labeling one feature resulted in an increase of its continuous information, whereas there was a reduction of continuous information on the non-labeled feature, but this varies for different item features. We concluded that these results point to trade-offs of how VWM capacity is allocated in dependence of labeling.

5. General Discussion

The first goal of this dissertation was to answer the question how consolidation in VWM operates. Specifically, I investigated this question by testing whether consolidation is a ballistic process in study 1. I found that enough time to consolidate is beneficial for VWM and that consolidation is not a ballistic process. Rather, attention needed for consolidation is subject to strategic choices of a person. Importantly, there was a consolidation time benefit even when labeling was inhibited through a suppression condition, ruling out the possibility that consolidation time is only used for verbal labeling (Experiment 1).

I then went on to answering the second goal whether labeling would facilitate consolidation. I found that labeling was indeed facilitating consolidation and was especially beneficial to the retention of continuous memory in VWM (study 2 and 3). However, there were also two limitations to this benefit. First, labeling only limitedly benefitted categorical VLTM (study 2) and second, labeling one item feature occasionally led to a cost to the non-labeled feature, by filtering this feature (especially color; study 3).

In the following, I will first briefly discuss the implications for consolidation in VWM and likewise, for verbal labeling. Finally, I address future directions to point out the open question from the present studies.

5.1. Implications of Consolidation During the Formation of a Memory Representation

In study 1, I showed that consolidation time is essential during the formation of a memory representation: it makes it more stable and thereby less susceptible to interference or time-based forgetting (Ricker & Cowan, 2014). In contrast, when consolidation time is sparse, only limited attention can be allocated towards the memory item. Study 1 showed that consolidation may appear to follow a ballistic manner, as the change of type of recall test (Experiment 1) or the fixation of focal attention (Experiment 2) was not sufficient to withdraw attention from the item currently being consolidated. However, if participants were forced to respond to a secondary task (Experiments 3+4), ongoing consolidation was interrupted. This finding is in line with the *resource-sharing models of attention allocation*. Thus, consolidation operates as follows: consolidation is a graded process that is subject to participants strategic choices of how attention is allocated to incoming information. This is an important finding as most recent research came to the conclusion that attention for consolidation induces an attentional bottleneck, supporting the view of an attentional-bottleneck model (Ricker & Hardman, 2017).

The finding of study 1 fit well with the eSTST model mentioned in the introduction. To recapitulate, in this model, incoming information (type) is activated and then bound to a token (consolidation), sustaining the activation of the representation in VWM. There are two important assumptions about the model in relation to consolidation: First, relevant information (type) can get an attentional excitatory boost, rendering it more likely to be bound to a token. Second, this binding (consolidation) can then suppress incoming information to sustain the consolidation of the representation. Thus, incoming information that interrupts ongoing consolidation needs to get a high attentional excitatory boost so that it can interrupt the consolidation of the former representation.

Following, a verbal label may facilitate consolidation leading to an excitatory boost of the type and further suppressing the reallocation of attention to new incoming information, which results in a more stable memory representation.

5.2. Are There Other Ways to Interrupt Consolidation?

To exemplify, consolidation was interrupted by a parity judgment task (Experiments 3 and 4). According to the eSTST model, the parity judgment task received an attentional boost, thereby interrupting the consolidation of the memory item. It remains an open question whether a distractor task requiring a secondary response is essential for this interruption. For example, a more difficult stimuli that is attentionally more demanding to consolidate may be sufficient to interrupt the consolidation process. Two findings suggest that this could be the case:

First, previous studies have shown, that task difficulty interacted with the degree of interference a distractor task had. Bayliss et al., (2015) showed that a more difficult task (reading an arithmetic problem or solving an arithmetic problem) interfered more with retrieval of letters. Similarly, Nieuwenstein and Wyble (2014) showed that a color discrimination task interfered more with recall than a color detection task. This suggests that consolidation can further be interrupted by the degree of difficulty of the new incoming information and it is open, whether this only needs to be a distractor task.

Second, for example, study 3 showed that the *color* feature was especially vulnerable to being filtered out. This suggests that color is especially costly and as a turn, requires more attention to form a representation. In study 1, I however only investigated the nature of consolidation of the orientation feature (visuo-spatial materials). If we assume that the consolidation of color is more costly, maybe the mere presentation of a new colored item during the consolidation of the first item is already incentive enough to reallocate attention to this new item? This would imply that consolidation of a colored memory item could already be interrupted under a serial recall procedure. However, it may also be the case that the mere presentation of color does not interrupt consolidation. Still, the two manipulations of changing type of recall test (Experiment 1 in study 1) and the locus of focal attention (Experiment 2 study 1) may not interrupt consolidation of a colored item.

5.3. Implications of Labeling During Consolidation

In Experiment 1 of study 1, we found that consolidation in VWM is different from verbal labeling, as there was a consolidation time benefit even under suppression. In study 2 and 3, we found that labeling the feature of a memory item during consolidation clearly improved VWM performance in contrast to suppression, in line with the hypothesis that labeling facilitates consolidation for VWM. In contrast, this benefit was only limitedly found for VLTM (study 2). This finding suggests that labeling may not facilitate long-term consolidation. In contrast, Ahmad et al. (2017) investigated the effect of presentation time in VLTM and found that recognition memory of pictures was better for longer presentation times (4000 ms) than shorter (1000 ms) ones. This implies that more time for consolidation benefits VLTM. Yet, labeling may only limitedly facilitate consolidation in VLTM. However, to fully answer this question whether labeling cannot assist long-term consolidation, the nature of consolidation and its time-course in VLTM first needs to be addressed in isolation.

For the case of VWM, study 2 and 3 revealed that labeling was in particular beneficial due to more continuous information or more continuous precision of the memory representation. This supports the hypothesis that a label activated categorical knowledge in VLTM and two visual memory traces are built: one stemming from the visual input and one from the activated category. The activated category can serve as a reference, meaning that information about the visual input can be added in relation to the category reference. This in turn, facilitates data compression, respectively the use of hierarchical representations (Brady et al., 2009; Brady & Alvarez, 2011, 2015a). As the memory trace no longer needs storage of all details of the memory representation, but just about the deviation of for example the color hue in relation to the category. This reduces memory load and leads to more efficiently form a memory representation.

There is one caveat about this hypothesis: it would have predicted a similar outcome for VLTM as shown in VWM, which was not the case in study 2. In contrast to VWM, labeling only limitedly benefitted VLTM when the color-shape binding was labeled, associated with an increase in categorical memory. This finding is predicted by the dual-trace hypothesis, yet which cannot explain the benefits for VWM. To recapitulate, the dual-trace hypothesis predicts that a verbal memory trace of the label and a visual trace of the input information is built, benefitting categorical memory. The finding that labeling differently affected VWM and VLTM suggests that there is a dissociation between the two memory systems. This was not predicted by any of the labeling hypotheses.

One possibility that could explain this dissociation is that the activation of categorical information in VLTM is also true for VLTM, but the representations are not translated to VLTM due to interference. This hypothesis also fits well with memory models assuming a tight link between VWM and VLTM, where VWM is an activated subset of VLTM. As mentioned earlier, VWM consist of a focus of attention holding relevant information and an activated part of VLTM where for example the label would activate VLTM knowledge. In these models, the activation of VLTM representations may not boost VLTM as the representations do not survive the proactive interference built up with learning more objects.

Another possibility is that maintaining continuous information in VLTM may be too costly. In line with this assumption, maintaining continuous representations in VLTM may not serve a practical purpose: VLTM serves to benefit VWM by activating existing representations (Oberauer, 2009). If we assume that VLTM would hold continuous representations, activating the correct association in VLTM may be very difficult, as it contains a lot of fine-tuned information. In contrast, storing only the category information in VLTM may be very beneficial, category information can be rapidly activated in VLTM and thereby benefit VWM. To exemplify, in our live, we see a lot of mugs varying in different colors, sizes and shapes. If a new mug is presented to us, we can immediately categorize it as a mug. This, even though the precise color of the mug does not match the VLTM representation of a mug. Hence, the detailed information about the color is not relevant.

In contrast to the beneficial effect of verbal labeling for the labeled feature, labeling during consolidation came at the expense of the non-labeled feature (see study 2 and 3).

This suggests that labeling can hamper or slow consolidation by suppressing the allocation of attention to the other item feature. The results showed that labeling the shape, orientation or frequency resulted in the involuntary filtering of continuous color, even though the color information was relevant to the task. This cost is in line with the hypothesis that a label directs attention to the labeled item: labeling the item allocates attention, what follows is that less attention can be directed towards the non-labeled item. As mentioned before, consolidating color information may require more attention than for example orientation. To exemplify, when the orientation of a colored triangle is labeled, less attention is directed to the non-labeled feature color. Thus, consolidating the color information on top of the orientation feature (including the orientation label) is too costly. In study 1, we found that attention is strategically allocated towards the item in order to consolidate it. In this case, a participant may think that consolidating the color in addition to the orientation feature is too costly, and attention is only directed to the orientation feature. In contrast, when the color is labeled, additionally consolidating the orientation feature does not require that much attention and the participant may be capable of holding both features in VWM.

5.4. Are There Other Ways to Facilitate Consolidation, Similarly as Labeling?

The aforementioned points directed on how labeling facilitates consolidation lead to the question whether other mechanism may also facilitate consolidation? For example, studies on prioritization have shown that priority instructions improved VWM performance (Allen & Ueno, 2018; Hu et al., 2014). For example, Allen and Ueno (2018) manipulated prioritization prior to the stimulus presentation by showing a display with the number of points as a reward for correct recall of each item: four points for the for the prioritized one and one point each for the remaining items. Then, four colored shapes were simultaneously presented, followed by a probe indicating which feature of one item is tested. Verbal recall performance of either the color or shape was better in the higher reward condition, suggesting that selectively directing attention to the higher reward item was beneficial for performance. This raises the possibility that prioritization may facilitate consolidation by directing attention to specific items.

6. Future Directions

There are two limitations of the present thesis. First, it is unclear what the exact mechanisms of the facilitation of labeling for consolidation are, for example whether labeling may speed up consolidation. Second, the role of attention in relation to the verbal labeling cost remains unclear. These two questions will be addressed in the following sections.

6.1. Can Labeling Speed Up Consolidation?

The present thesis is limited in regard to answering the question how verbal labels can facilitate consolidation. One possibility is that labeling speeds up consolidation. To answer this question, a new experiment is needed in which consolidation and labeling are directly manipulated. I would suggest the following procedure: In line with study 1, four visuo-spatial memory items will be sequentially presented, followed by a mask to overwrite sensory memory. Then, a free time interval follows. To more directly assess how consolidation and labeling may interact, I suggest to use more consolidation time intervals, for example: 200 ms, 400 ms, 600 ms, 800 ms, 1000 ms, 1200 ms, 1600 ms, 2000 ms to better detect a pattern across time. The challenge is that the way labeling was manipulated before will no longer work in assessing an effect on consolidation. In the previous studies, participants were given 1000 ms to label the item. Ricker and Hardman (2017) showed that consolidation of these visuo-spatial items takes about 800ms, which would already be sufficient time to consolidate the item and a labeling effect may no longer be detectable. Thus, I suggest to either present a written label of the item (e.g. right, bottom, left, top); or play a sound during the consolidation interval. The efficacy of these manipulations as well as the consolidation time intervals need to be tested in a pilot study. As before, the labeling condition will be contrasted to a suppression condition. Then, participants will be tested on all memory items in presumably serial order.

The results should show flatter curves across the serial input positions for all consolidation times in contrast to suppression, suggesting that labeling speeds up consolidation. A possible outcome is that labeling may further be more beneficial for shorter consolidation times than longer ones, suggesting that the two processes interact. In any case, enough time for consolidation in combination with the verbal label should result in the most robust representation and most sufficiently counteracting the capacity limitations in VWM.

6.2. What Role Does Attention Play for Labeling?

Another open question is what role attention plays during labeling. This question emerged in particular in study 3, where we found support for the hypothesis that a label directs attention to the labeled item, which then came at a cost for the non-labeled item that was filtered as a result. However, this study cannot fully distinguish whether this cost emerged due to the label or whether it was due to mere attention. This requires another experiment that directly manipulates attention and labeling during the free time interval. A possible experiment could look as follows: a single memory item consisting of two features (e.g. colored triangle, as used in study 3) will be presented in the middle of the screen. The single item will make the task attentionally less demanding. Then, a free item interval will either be used for a labeling condition (as in study 3) or a secondary task requiring attention. In the labeling condition, participants will either label the color or the orientation. In the distractor task condition, participants will be asked to do a color matching task. For this purpose, a colored dot will be presented in the middle of the screen, with a grey dot next to it and both will be surrounded by a color wheel. Participants have to match the grey probe to the color of the colored dot as accurately as possible. Then, participants will be tested on both features of the memory item.

In the labeling condition, I expect more continuous color and orientation for color and orientation labeling, respectively. More important to answer the question is the finding of a cost for continuous color with orientation labeling. If this cost will also emerge in the attention condition, this would speak for an effect of attention in the filtering process. In this case, the item feature was not filtered due to the verbal label, but due to less attention being directed towards this feature.

7. Conclusion

The aim of this thesis was first to investigate how consolidation operates and second whether labeling could be used to facilitate consolidation. I found that consolidation does not operate in a ballistic manner and that attention can be strategically allocated during the formation of a memory representation. Furthermore, I found that verbal labeling can facilitate consolidation, and this benefit was especially due to more continuous information in VWM, but there are two limitations. First, the short-term labeling benefit only limitedly translated into a long-term benefit. Second, labeling one item feature also occasionally resulted in a cost for the non-labeled item feature.

PART II - EMPIRICAL STUDIES

8. Consolidation is not a ballistic process: Evidence that a distractor task disrupts ongoing consolidation in visual working memory

Clara Overkott and Alessandra S. Souza

University of Zurich

Status: *under review* in Journal of Experimental Psychology: General (25.06.2020)

Author's contribution

Conceptualization, methodology, and software C.O. and A.S.S.; investigation, formal analysis, visualization, data curation, and writing – original draft preparation, C.O.; writing – review and editing, supervision, project administration, resources, and funding acquisition, A.S.S.

Acknowledgments

This research was support by a grant from the Swiss National Science Foundation (100019_169302) awarded to A. S. Souza. We thank Jasmine Kerr and Vanessa Vallesi for their assistance during data collection. We further thank all students taking part in the ExPra FS2019 seminar (Introductory Course on Research Methods in Experimental Psychology) of the Cognitive Psychology Unit of the University for their feedback on the experimental set-up and their assistance during the data collection.

8.1. Abstract

Short-term consolidation is an attentionally-demanding and time-consuming process that creates stable working memory representations. Consolidation has been assumed to occur in an all-or-none fashion, following a ballistic process that cannot be interrupted once started. Evidence for this claim comes from the finding that, for sequentially presented items for a serial recall test, consolidation of the first-presented item is not interrupted when the next memory item is presented, creating an attentional blink. Across four preregistered experiments, we examined whether consolidation is interrupted if (a) recall order is unpredictable (Experiment 1), (b) encoding does not require shifts of visuospatial attention (Experiment 2), (c) a distractor task follows each item (Experiments 3 and 4), and (d) items are encoded simultaneously (Experiment 4). All-or-none consolidation of the first-presented item was found irrespective of recall order and shifts of visuospatial attention. Nevertheless, presenting a distractor task interrupted ongoing consolidation. This finding challenges the assumption that consolidation in working memory is a ballistic process. It shows that the first item in a sequence is prioritized: attention lingers on it allowing it to be fully consolidated, but this is not an immutable feature of the short-term consolidation of a memory representation.

8.2. Introduction

Every moment, we are inundated with visual information. This rich perceptual experience stands in contrast to how little we can retain in our mind about our visual surroundings a moment later. This is because the ability to hold visual information in mind is limited by the capacity of visual working memory: Only a handful of representations can be maintained simultaneously in this system for ongoing processing

(Cowan, 2017; Oberauer et al., 2016; Oberauer & Hein, 2012). In order for representations to be retained in working memory, the traces created from sensory inputs first need to be transformed into a stable memory representation. This is achieved through the process of short-term consolidation (Bayliss et al., 2015; Chun & Potter, 1995; Jolicœur & Dell'Acqua, 1998; Ricker, 2015; Ricker et al., 2018; Ricker & Cowan, 2014).

One line of research has proposed that consolidation is an attention-demanding process that occurs in a ballistic manner, meaning that once sensory information is being translated into a stable working memory representation, this process cannot be stopped (Jolicœur & Dell'Acqua, 1998; Stevanovski & Jolicœur, 2007; Tombu et al., 2011). According to this view, consolidation functions in an all-or-none manner: either representations are consolidated and accessible in working memory, or they are lost. This implies that as information is being consolidated, any other cognitive activity requiring attention has to wait. Evidence for this claim is still limited, and another line of research has provided evidence consistent with consolidation being a graded process that can lead to more or less consolidated representations (Bayliss et al., 2015; Kandemir et al., 2017; Nieuwenstein & Wyble, 2014). What is the source of these divergent views on consolidation? Here we put the assumption that consolidation occurs in an all-or-none fashion to a stronger test by contrasting the features of the paradigms used in these two research lines. This allowed us to demonstrate that consolidation appears to occur in an all-or-none matter, not because of structural constraints, but because of functional ones: participants do not have enough incentive to stop ongoing consolidation.

In the following, we will review what is currently understood as consolidation, and the evidence bearing claims that it proceeds in a ballistic way. Finally, we will present our hypotheses and the manipulations implemented across experiments.

8.2.1. Short-Term Consolidation

Short-term consolidation, or simply consolidation, is currently understood as an attentionally demanding process (Bayliss et al., 2015; De Schrijver & Barrouillet, 2017; Jolicœur & Dell'Acqua, 1998; Ricker et al., 2018; Stevanovski & Jolicœur, 2007). When attention is directed to a sensory trace, this trace is consolidated into a stable working memory representation that can be accessed later on. Without consolidation, no information about the sensory trace remains in memory after it is overwritten by a new input or it fades away. Consolidation has been differentiated from perceptual encoding: while encoding depends directly on the sensory trace and is assumed to end with the presentation of a mask, consolidation continues during the period after the mask as long as attention is allowed to dwell on this representation (Nieuwenstein & Wyble, 2014; Ricker & Cowan, 2014; Ricker & Hardman, 2017).

Consolidation can be differentiated from other working memory processes, such as rehearsal (Bayliss et al., 2015; De Schrijver & Barrouillet, 2017) and refreshing (Bayliss et al., 2015; but: De Schrijver & Barrouillet, 2017; Vergauwe et al., 2019) that presumably occur after consolidation, e.g. well within during the maintenance phase. Rehearsal consists of the covert or overt articulation of verbal information, whereas refreshing involves focusing attention on working memory representations to boost them. Critically, consolidation takes place in the short period after the last memory item was presented, whereas these other processes can be initiated any time within the maintenance interval and additionally, they can operate over all memory representations.

Consolidation is also assumed to be a time-consuming process. Evidence for this comes from the finding that memory improves when more time is provided for the consolidation of an item shortly after it has been presented. For example, Ricker and Hardman (2017) investigated the time course of consolidation of simple stimuli (e.g., the position of a dot at the edge of a ring) varying along a continuous scale. They sequentially presented four items for a brief interval (150 ms), followed shortly by a visual mask (50 ms). Critically, a blank screen was inserted between the offset of the mask and the onset of the next stimulus. By varying the duration of this blank interval between 17 ms and 2167 ms, they observed that performance improved monotonically to up to 800 ms and remained stable thereafter.

So far, the exact time-course of consolidation is still not clear. Nieuwenstein and Wyble (2014) observed that consolidation was completed for simultaneously presented letters or Chinese symbols within 1000 ms. Earlier work by Jolicœur and Dell'Acqua (1998) and Stevanovski and Jolicœur (2007) suggested that consolidation of verbal items took between 500 and 1600 ms. In contrast, increases in consolidation time remained beneficial up to 3100 ms in a complex span tasks for verbal materials, which involves the presentation of a distractor task following each memory item (Bayliss et al., 2015; De Schrijver & Barrouillet, 2017). Overall, the time needed to consolidate information seems to depend on the memory materials and the difficulty of the task (Ricker et al., 2018).

8.2.2. Consolidation: A Ballistic Process?

Consolidation in working memory has been described as a ballistic process, which means that consolidation occurs in an all-or-none manner (Ricker, 2015; Ricker & Hardman, 2017). This view is based on the strict *attentional bottleneck model* proposed by Jolicœur and Dell'Acqua (1998). They assume that the attentional resource needed for consolidation is capacity-limited and can only proceed in serial manner, thereby postponing any following task requiring attention. This model is supported by studies finding that the presentation of a secondary task shortly after the memoranda increased reaction times (RT) on this secondary task (Jolicœur & Dell'Acqua, 1998; Stevanovski & Jolicœur, 2007; Tombu et al., 2011). This was interpreted as indicating that while the memoranda were being consolidated, attention could not be reallocated to the secondary task, thereby postponing its processing.

This implies that once consolidation of a visual input has started, it has to finish before attention can be directed towards something new. It is well established that when people engage with the processing of one piece of information (e.g., a picture), they often cannot report the identity of new, subsequently presented information (within 200-500 ms thereafter). This phenomenon is known as the attentional blink (Chun & Potter, 1995; Shapiro et al., 1997; Wyble et al., 2009, 2011, 2015). Attentional blink theories claim that this blink is induced by consolidation: the creation of a stable representation of the first target blocks attention, and during this time, the representation of the second target is overwritten without the chance of being consolidated (Ricker & Hardman, 2017; Wyble et al., 2009, 2011, 2015).

Ricker and Hardman (2017) assessed whether the attentional blink effect is observed in a visual working memory task. They sequentially presented four visual memory items and varied their temporal separation (200 ms vs. 2000 ms) across trials. They observed that longer consolidation time was beneficial for items 2, 3, and 4 in the sequence, but not for item 1 which was always recalled best and with equal accuracy irrespective of consolidation time. From now on, we will refer to this observation as the *first-item effect*. They interpreted this finding as an indication that the first item was always consolidated: once consolidation of this representation started, it was not interrupted by the presentation of the second item. Instead, increasing the consolidation time after item 1 improved memory for item 2, because it reduced the likelihood of item 2 being blinked. This led them to conclude that consolidation in visual working memory is therefore a ballistic process: consolidation of an item once started cannot be stopped.

8.2.3. Challenging Views on the Ballistic Nature of Consolidation

There is evidence challenging the assumption that consolidation runs in an all-or-none manner (Bayliss et al., 2015; Kandemir et al., 2017; Nieuwenstein & Wyble, 2014). This assumption supports *resource-sharing models of central attention allocation*, which are in contrast to the attentional bottleneck model (Lehle & Hübner, 2009; Tombu & Jolicœur, 2003). Similar to strict bottleneck models, resource-sharing models assume that attention is a limited resource, but unlike the former attention is assumed to be distributed across the incoming information by allowing parallel processing. In other words, attention can be allocated towards multiple incoming information and is not restricted to only one at a time.

Such a resource-sharing model gained support in, for example, Nieuwenstein and Wyble (2014) who asked participants to retain a simultaneously presented array of 4 letters for a recall test. In half of the trials, participants responded to a parity judgment task during the retention interval of the memory task. Critically, the temporal separation between the offset of the memory array and the presentation of the distractor task was set to 50, 300, or 800 ms. If consolidation of the memory array cannot be interrupted, memory performance should remain constant regardless of the temporal separation to the distractor task. Performance of the distractor task should be impaired when it is presented closer in time with the memory array, because it has to be postponed until memory consolidation is finished. If consolidation of the memory array is interrupted by the distractor task, memory performance should be worst when the distractor task occurs shortly after the memoranda (because the time for consolidation is short) but improve with longer memory-distractor separation (because consolidation time is long). In accordance with the latter, the distractor task interfered with memory performance, and this interference was largest at the shortest interval, and smallest at the longest interval. This result suggests that consolidation can be interrupted if a distractor task is presented after a memory item. Moreover, the interference effect depended on the type of secondary task: a color discrimination task produced a stronger interference effect than a color detection task. This is expected because discrimination tasks involve response selection, whereas detection does not, which consumes more central attention.

Similar results were obtained by Kandemir et al. (2017), who found that a distractor task (judge whether a picture contained a red or green frame) interfered with memory recall more after a short compared to a longer consolidation time. This interference was not present when the same pictures were presented during maintenance without the decision task. Likewise, Bayliss et al. (2015) found that a distractor task (arithmetic problem) interfered with memory performance for letters, and this interference was larger for shorter consolidation intervals than longer ones. The interference effect further depended on the difficulty of the distractor task, arguably because the more difficult task required more central attention.

There are some caveats though. Nieuwenstein and Wyble (2014) and Kandemir et al., (2017) presented four letters simultaneously, and participants recalled all items at the end. It is unclear whether the simultaneously presented items were consolidated in parallel or serially. If they were consolidated serially, and assuming a left-to-right reading bias, it might still be the case that the left-most letter was encoded first in a ballistic fashion, followed by the second, and so forth, until consolidation of one of the serially encoded letters was prevented by the distractor task. This assumption remains untested because recall was scored as the average across all letters. Hence, it is possible that the first letter was always recalled correctly irrespective of the time separation with the distractor task. If this is the case, there would be no disagreement between the results of Nieuwenstein and Wyble (2014), Kandemir et al., (2017) and Ricker and Hardman (2017): consolidation is a ballistic process, but methodological differences between the experiments led them to different conclusions about the nature of this process. This caveat is less of a concern in the study of Bayliss et al. (2015), in which the letters were sequentially presented, but recall was scored as average performance of the presented letters nevertheless.

8.2.4. Stronger Test of the Nature of Consolidation?

The observation of the *first-item effect* – e.g., the fact that the first item was recalled best irrespective of consolidation time – led Ricker and Hardman (2017) to propose that consolidation cannot be interrupted once it has started, therefore yielding an attentional blink. Although they observed the first-item effect across several experiments, the authors never considered whether this effect could be the result of a strategic choice rather than a structural feature of the consolidation process. Strategic allocation of capacity limitations is perfectly in line with resource-sharing models of central attention allocation.

There are reasons to speculate that their experimental set-up could have favored the observation of a first-item effect due to strategic preferences. First, Ricker and Hardman (2017) asked participants to recall all memory items in serial order. This requirement could have led participants to prioritize the consolidation of the first item given that this would allow them to at least recall one item with high accuracy. Second, they did not implement an articulatory suppression procedure. By varying the consolidation time, participants were also allowed more time to label the visual items. Verbal labeling substantially improves recall in continuous reproduction tasks (Overkott & Souza, 2020; Souza et al., 2020; Souza & Skóra, 2017). Critically, participants may have started labeling the first item but they might have not been able to label the remaining items when the consolidation time was short. This would create a situation in which only the first item is labeled irrespective of consolidation time, thereby explaining the appearance of the first-item effect.

Third, another feature of their procedure refers to how memory items were presented. Ricker and Hardman (2017) presented their memory items in different and unpredictable locations around the center of the screen (4 items presented in random subset of 8 locations). Hence, participants were obliged to move their visual attention to unpredictable locations. If participants were strategically focusing all their attention to the first presented item, they could have chosen to not move their visual attention to the next upcoming item while they were not finished with consolidation. In this scenario, participants would consolidate the first item irrespectively of consolidation time, leading to a first-item effect.

Fourth, Ricker and Hardman (2017) only investigated whether consolidation could be interrupted by the presentation of a subsequent memory item. In contrast, Nieuwenstein and Wyble (2014) presented a distractor task after the memory array, and observed that this task interfered with the formation of the memory trace when it appeared shortly after the memoranda. The parity judgment task requires central attention to be processed and it demands an immediate response. It is possible that the presentation of another memory item could not be a sufficiently motivating event to draw attention away from the previously presented memoranda, thereby stopping its consolidation. In contrast, the presentation of a distractor task demanding an immediate response may force withdrawal of attention away from the memoranda.

8.2.5. The Present Study

The aim of the present study was to test whether the first-item effect truly reflects the operation of an all-or-none process, or whether it could be an artifact of the experimental procedure. Answering this question can shed light on the true nature of consolidation: does it operate in an all-or-none fashion, as proposed by strict bottleneck accounts, or is it a graded process under strategic control? The latter could explain why, under some circumstances, consolidation may appear to operate in an all-or-none fashion, whereas under others it occurs in a graded manner.

Across four preregistered experiments, we systematically tested whether the first-item effect would vanish when: (a) verbal labeling was suppressed with an articulatory suppression procedure (Exp. 1), (b) the first-presented item was not recalled first (Exp. 1), (c) encoding uncertainty was removed (Exp. 2), (d) shifts of spatial attention were not required (Exp. 2), and (e) when an attention-demanding distractor task was inserted after the memory item (Exp. 3. & 4), and whether this is modulated by mode of presentation of the memoranda (e.g., simultaneously or sequentially; Exp. 4). Table 1 presents an overview of the manipulations implemented across our experiments.

To foreshadow our results, we found that (a) verbal labeling could not explain the consolidation time effect. Moreover, consolidation was not interrupted when (b) the first-presented item was not recalled first, nor when (c) the encoding uncertainty was removed, nor (d) when shifts of spatial attention were not required. In contrast, consolidation was interrupted when (e) an attention-demanding distractor task was inserted in between the memory item and the memory recall. These results support the notion that consolidation is a graded process under strategic control – in line with the resource-sharing models and against attentional bottleneck models of central attention.

Table 1

Overview of the Features of Experiments 1-4

Exp.	Condition Name	Mode of Presentation	Memoranda Position	Recall Test	Distractor Task	Suppression
1	Serial	Sequential	4 out of 8 locations	***Serial***	No	Yes
	Random	Sequential	4 out of 8 locations	***Random***	No	Yes
2	Periphery	Sequential	***4 out 4 locations***	Serial	No	No
	Center	Sequential	***1 (center) location***	Serial	No	No
3	No-distractor	Sequential	4 fixed locations	Serial	***No***	No
	Distractor	Sequential	4 fixed locations	Serial	***Yes***	No
4	Distractor + Sequential	***Sequential***	4 fixed locations	Serial	Yes	No
	Distractor + Simultaneous	***Simultaneous***	4 fixed locations	Serial	Yes	No

Note. The critical manipulation between conditions in the experiment is presented in bold, italics font.

8.3. Experiment 1

Experiment 1 assessed whether the requirement to recall all items in forward serial order could explain the first-item effect. In a forward serial-order test, participants are aware of the order in which items will be tested. In this situation, participants may selectively choose to keep consolidating the first-presented item because they see no incentive in reducing their recall performance for the very first tested item.

In a random recall test procedure, in contrast, every item is equally likely to be tested first, second, and so forth. Hence, optimal preparation for the test would not require having the first item ready to be recalled. This could arguably motivate participants to give equal priority to all items. If the first-item effect arises from preparation for the recall test, exposing participants to a random recall test may remove this incentive, leading to the observation of a consolidation time benefit for all serial input positions.

To test for this possibility, the present experiment varied the type of recall test across two different experimental sessions. In one session, participants recalled items in forward serial order, replicating the procedure of Ricker and Hardman (2017). In another session, recall of the four items was probed in random order. Across these two conditions we were able to address two further confounds.

8.3.1. Confound of Output Position

One issue in the serial recall procedure in Ricker and Hardman (2017) is that it confounds input and output position effects. In any memory test, the first recalled item is immune to output interference. For forward serial recall, this is always the first presented item. Consolidation time may be useful to stabilize a memory trace to survive output interference, and this may explain why the first item showed no effect of consolidation time.

In contrast, a random recall procedure guarantees that all items are tested in conditions of reduced output interference (e.g., the first recalled item), and with output interference (e.g., for subsequent recalled items). Hence, if the finding in Ricker and Hardman (2017) could be explained as an effect of output position, all items tested first in the random recall test should show no consolidation effect, whereas consolidation benefits should be observed for all items tested in the subsequent output positions. This finding would support the conclusion that consolidation time serves to stabilize a working memory representation and make it more robust to interference.

8.3.2. Confound of Verbal Labels

Ricker and Hardman (2017) did not test for the possibility that the beneficial effect of consolidation time was due to verbal labels. There is a possibility that participants did not use the no-distractor interval for consolidation, but instead used it to verbally label the memoranda. Souza and Skóra (2017) recently showed that a 1-s interval following each item provides enough time to yield a verbal labeling benefit. In their studies, participants were asked to label the colors of four sequentially presented disks, and later recall the colors of all items using a continuous color wheel. There was a clear labeling benefit in contrast to a verbal suppression condition where participants repeatedly said “ba ba ba” aloud. Thus, there is a possibility that the reported consolidation time benefit may not be a consequence of consolidation time, but rather a verbal labeling benefit. To prohibit the possibility that participants were benefitting from verbal labels, participants were asked to say “ba ba ba” aloud during the item presentation phase.

It is worth noting that Bayliss et al. (2015) showed that the consolidation of a series of letters was unaffected by verbal suppression. However, in their study, participants labeled the letters aloud during their presentation, and the suppression procedure was in effect during the blank interval after the series of letters were presented. Furthermore, their materials were verbal in nature, and hence verbal labeling (or rehearsal) might not add much to their representation. For visual materials, in contrast, verbal labels could be much more beneficial.

This experiment was preregistered on the Open Science Framework, OSF (https://osf.io/g94u5/). Data, materials, and analysis scripts are openly accessible at https://osf.io/gncbq/.

8.3.3. Method

8.3.3.1. Participants

For all experiments reported here, only participants with German (or Swiss-German) mother tongue or very good German skills, aged between 18-35 years and reporting normal color vision and normal or corrected-to-normal visual acuity were eligible to take part in the studies. Participants signed an informed consent prior to the study and were debriefed at the end. The experimental protocol in each experiment was in accordance with the Institutional Review Board of the Psychology Institute from the University of Zurich, and it did not require special ethical approval.

Fifty-two (M = 23.83 years old; SD = 3.31; 37 women) students of the University of Zurich took part in Experiment 1. Two participants were excluded, one due to an experimental error (error-coding of participant ID) and one because of prior participation in a similar experiment. In total, the data of 50 participants were retained for analysis. We started collecting data of 30 participants. As stated in the preregistration, we aimed to report BFs ≥ 10 and that we aimed to collect up to 60 participants to match this criterion.

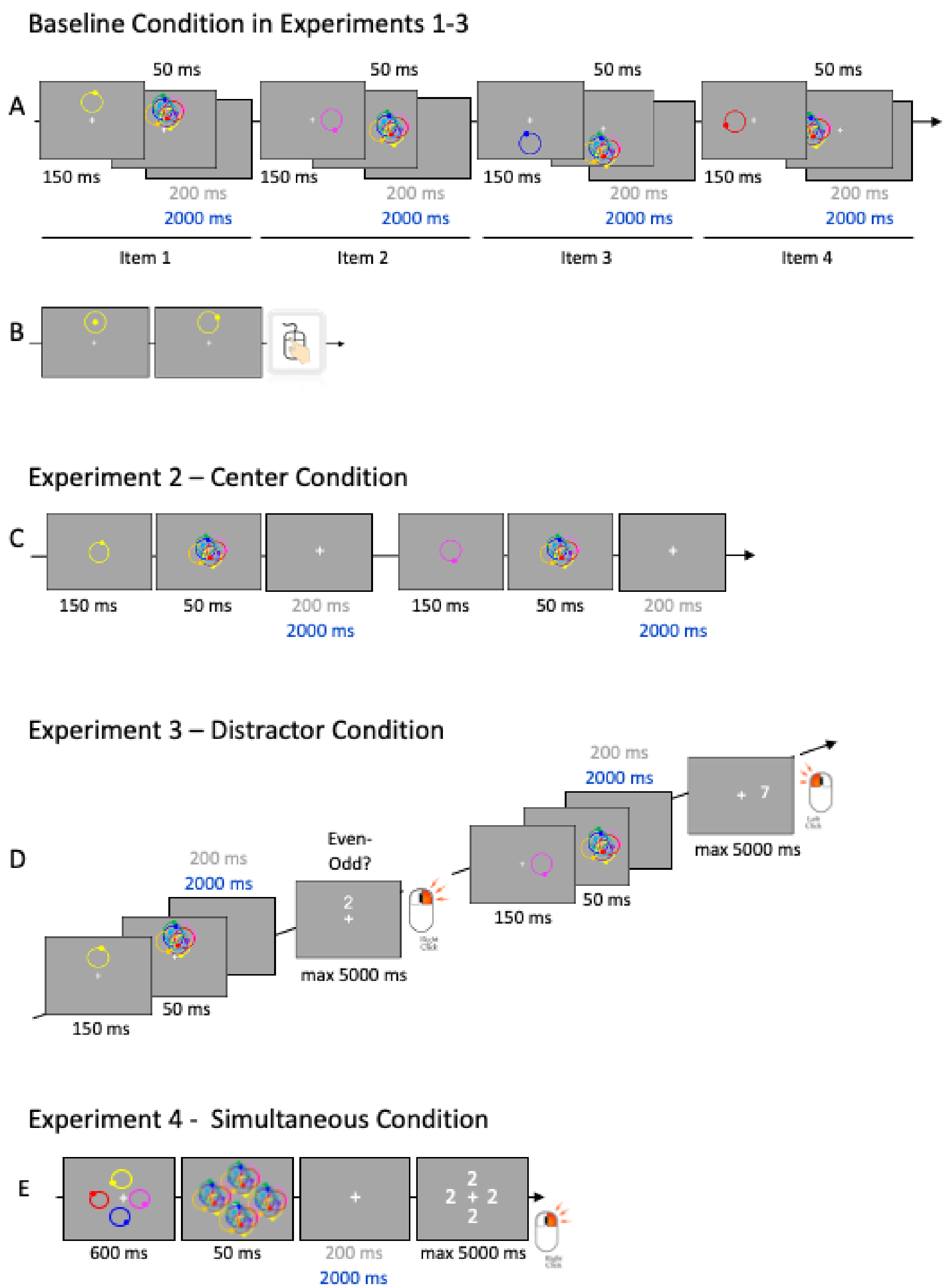
8.3.3.2. Materials and Procedure

The task was programmed in Psychophysics toolbox (Brainard, 1997; Pelli, 1997) running in Matlab. The procedure of Experiment 1 was modeled after the Experiment 1 of Ricker and Hardman (2017), as illustrated in Figure 1A. In the study phase, four memory items (e.g., a ring with a dot at the edge) were sequentially presented, each followed by a mask, and a blank consolidation interval whose duration was either short (200 ms) or long (2000 ms). The mask was used to stop ongoing encoding of the memory item and to verify that after the mask, only consolidation was measured (Ricker, 2015; Ricker & Sandry, 2018). Afterwards in the memory test phase, memory for the studied items was tested in serial or random order (test phase; see Figure 1B). Participants completed two sessions. In one session, recall of all items was probed in serial order. In the other session, recall of all items was probed in random order. Session order was counterbalanced across participants. To inhibit verbal labeling, participants repeated “ba ba ba” throughout the study phase and their verbalizations were recorded.

At the start of each trial, a fixation cross appeared and remained on screen throughout the whole trial procedure (e.g., study and test phase). The first memory item appeared 500 ms thereafter. Each item was presented for 150 ms. The item consisted of a thin colored circular ring (diameter = 2.5 cm) with a dot (diameter = 0.4 cm) at its edge. The angular locations of the dots were selected randomly for every item between 1 and 360 degrees. The task was to memorize the precise angular location of the dot on the ring. Each ring was presented in one out of eight different spatial positions on the screen, with the constraint that each position was used only once in a trial. The eight ring positions were evenly distributed along an imaginary circle centered in the middle of the screen (radius = 4.5 cm). Furthermore, each of the eight item locations was associated with an invariant color across the experiment. The following color values were assigned: yellow (RGB = 255 255 0), purple (RGB = 255 0 128), green (RGB = 0 255 0), red (RGB = 255 0 0), turquoise (RGB = 90 190 255), violet (RGB = 138 20 236), orange (RGB = 255 90 0) and blue (RGB = 0 0 255).

Figure 1

Flow of Evens of Each Experiment



Note. Panel A shows the item presentation phase and Panel B the test phase that was used in the baseline condition of Experiments 1-3. Panel C gives an example of the central-presentation condition in Experiment 2. Panel D gives an example of the distractor task condition in Experiment 3, wherein participants judged whether a number was odd or even after every item presentation. Panel E shows the simultaneous presentation (+ distractor task) condition of Experiment 4.

A mask was presented for 50 ms after every item. The mask consisted of eight colored rings (memory item colors) with randomly allocated dots, with the center of each ring slightly shifted from the center of the target ring. Then, a blank screen in which only the fixation cross was visible appeared. The duration of this blank screen was set to either 200 ms or 2000 ms (hereafter the consolidation time). Consolidation time was manipulated between trials, and thus remained the same for the four items within a trial. In total, 8 practice trials and 4 blocks of 30 experimental trials were completed in each session.

For the forward serial order recall session, the order of testing of the memoranda was the same as the presentation order. For the random order recall session, the order of testing was randomly determined in each trial with the constraint that each serial input position was tested an equal amount of times in each output (testing) position across all trials of this condition. During the test phase, a colored ring with a colored dot in its center was presented in the original location of the studied item (see Figure 1B). Participants were instructed to reconstruct the precise location of the dot in the tested ring by moving the dot (using the mouse) to the corresponding angle on the edge of the ring. Once they were satisfied with the selected angular location, they were asked to confirm their response with a left mouse-button click. This procedure was repeated until all studied items were tested. At the end of the trial, participants received feedback on the accuracy of their responses. During the feedback screen, the selected location of each dot and the correct location of the dot (indicated by a white dot on the ring) were simultaneously displayed onscreen for 3000 ms. Before each trial, participants were first reminded to say "ba ba ba" aloud to inhibit any kind of verbal labeling. Participants then initiated each trial in a self-paced manner by a press on the spacebar.

8.3.4. Data Analysis

The data was first submitted to Bayesian ANOVAs (BANOVA) and Bayesian *t*-tests. Bayesian inference has several statistical advantages over the commonly used frequentist ANOVA approach, where *p*-values are reported. For example, the *p*-value has a tendency to overestimate the evidence in favor of the alternative hypothesis (Wetzels et al., 2011). For Bayesian analysis, commonly Bayes Factors (BFs) are provided. A BF is quantified by the evidence for the one hypothesis (e.g. alternative) against the other (e.g. null), given the observed data. An advantage is that a BF provides the strength of the evidence for either the null or the alternative hypothesis. In other words, the BF can be reported in favor of the alternative (BF_{10}) or the null (BF_{01}), where $BF_{01} = (1/BF_{10})$. A BF_{10} larger than 1 yields evidence *for*, and a BF_{10} lower than 1 evidence *against* an effect. A BF_{10} of 10 indicates that the alternative hypothesis is 10 times more likely than the null. Usually, BFs > 3 are regarded as providing substantial evidence for one hypothesis over the other. BFs were computed in line with Rouder et al. (2012) by using the BayesFactor package (Morey & Rouder, 2015) which is implemented in R (R Core Team, 2014).

To compute the BFs, a priori beliefs of the Cauchy distribution about the probability of an effect must be chosen (Rouder et al., 2012). In this study, the most conservative default prior with a scaling factor of $(\sqrt{2})/2$ which is provided by the package was used. According to Rouder et al. (2017) the prior specification matters and must be judiciously chosen, but the prior does not substantially change the evidence when specified in the range of 0.2 and 1.

8.3.5. Results

Our dependent variable was recall error, which is the absolute distance between the true angle of the memory dot and the response of the participant. This value can range between 0 and 180 degrees, with lower values indicating better performance. Recall error in Experiment 1 is presented in Figure 2. Figure 2A and 2B present performance across the four serial input positions and consolidation time for the serial recall and the random recall conditions, respectively. Figure 2C shows the recall error in the random recall condition as a function of the output position.

To recapitulate, the first-item effect is the finding that a longer consolidation time does not improve recall of the first-presented item in contrast to shorter consolidation time. In contrast, the subsequent serial input positions show evidence of a consolidation time benefit, meaning that more consolidation time results in lower recall error. For all experiments, we will first state the evidence for or against the observation of a first-item effect. Then, we will report the remaining analyses concerning the full pattern of the data.

8.3.5.1. First-Item Effect

Figures 2A and 2B shows that recall for the first-item tended to be worse with longer consolidation time. This pattern is reversed for the remaining serial positions. To estimate the effect of consolidation time across the two recall test conditions for the first-presented item, we conducted a 2-way BANOVA having consolidation time (200 ms vs. 2000 ms) and recall test (serial vs. random) as fixed-, and subject as random-predictor.

The best BANOVA model included the two main effects of consolidation time and recall test. This model was substantially favored against the model including the interaction (see Table 2). The inclusion of the consolidation time effect into the model reflects the fact that performance with long consolidation time was worse than with short consolidation. This pattern can also be detected in the data reported by Ricker and Hardman (2017). This finding is opposed to what would be expected under the assumption that consolidation time is beneficial for the first input position.

To gauge the evidence against a beneficial effect of consolidation time on the first-item, we ran two one-sided Bayesian *t*-tests which directly tested the hypothesis that shorter consolidation time should result in worse performance than longer consolidation time. Table 3 shows the evidence for a consolidation time benefit on the first-item across all experiments. For both the serial and the random recall conditions in Experiment 1, there was substantial evidence against a consolidation time benefit. Hence, the manipulation of the recall requirements did not change the fact that the first item was always consolidated irrespective of the consolidation time.

To visualize the first item effect more specifically, we present the posteriors of the consolidation time effect on the first item for the serial and the random recall test in Figure 3A. The posteriors were drawn from the full model that included both main effects (consolidation and recall test) and their interaction. To estimate the first-item effect, we subtracted the posterior estimates of the long consolidation condition with the posteriors of the short condition for each recall test. Each posterior is presented with its mean (*M*) and indication of its highest density interval (HDI) represented by the bar underneath the curve. The HDI reflects the range of values that covers 95% of the posterior. As shown in

Figure 3A, the mean of the posterior was positive for both conditions indicating that a longer consolidation interval was associated with an increase in the recall error for both recall tests. The overlap in the HDIs of the two conditions indicates that the effect of consolidation time was similar for both recall tests.

Table 2

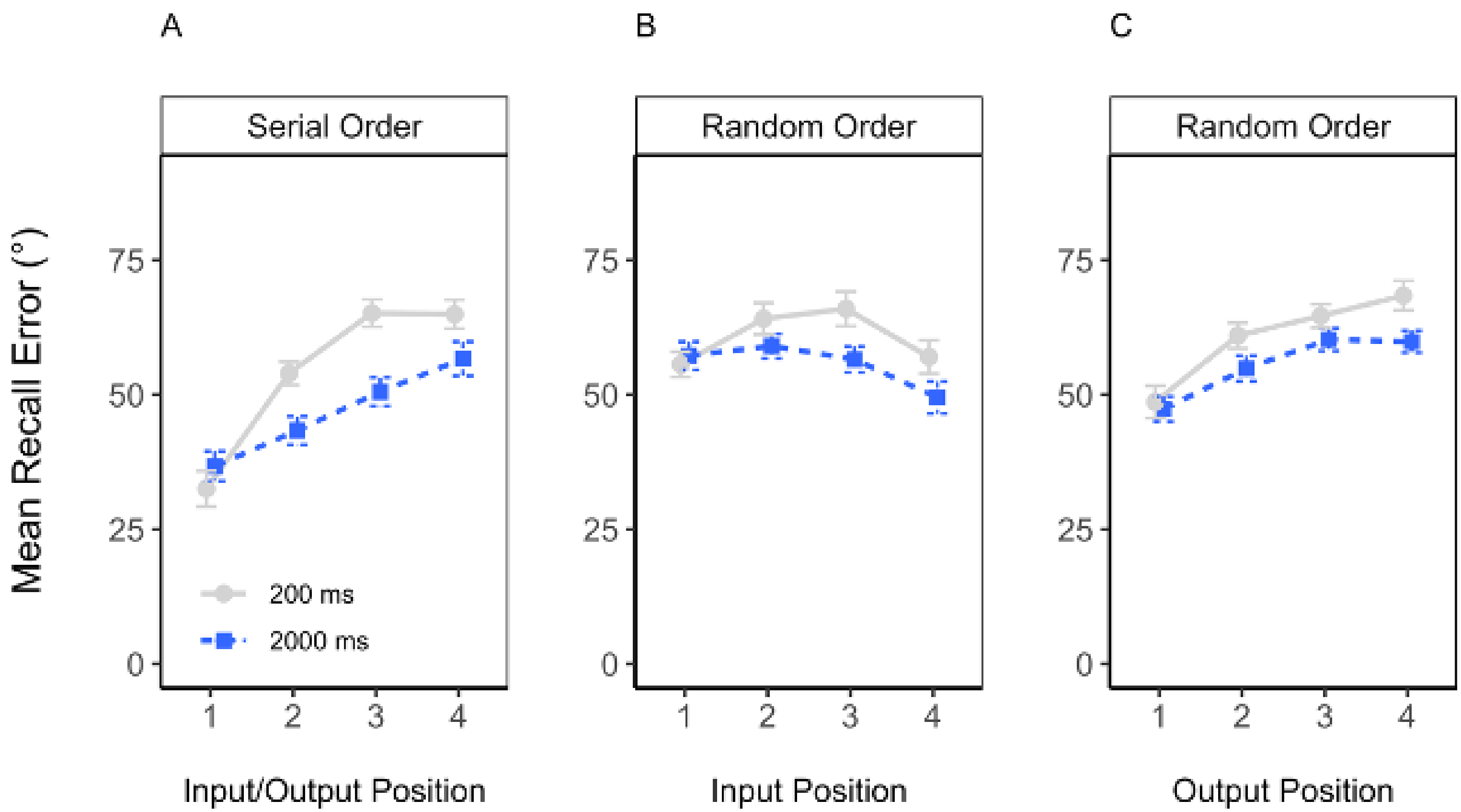
Bayes Factor (BF) of Models with Different Fixed Effects Over the Null and BF favoring the Best Model Over the Model Specified in Each Row (BF_{Best}/BF_{Mrow}) for Recall of the First-Item in Experiment 1

		Included Fixed Effects				
Condition	Model n°	Consolidation	Recall Test	Cons. x Test	BF_{10}	BF_{Best}/BF_{Mrow}
Input Position 1	1	✓	✓	✓	8.81×10^{153}	9.47
	2	✓	✓	**---**	$\mathbf{8.35 \times 10^{154}}$	**1**
	3	✓	---	---	9.48	8.81×10^{153}
	4	---	✓	---	6.10×10^{153}	13.68

Note. ✓ = effect included in the model. Best model is printed in bold. Best model = model with higher BF over the Null.

Figure 2

Recall Error as a Function of Serial Input Position and Output Order for the Serial and Random Recall Tests in Experiment 1



Note. Error bars represent the 95% within-subjects confidence interval.

Table 3

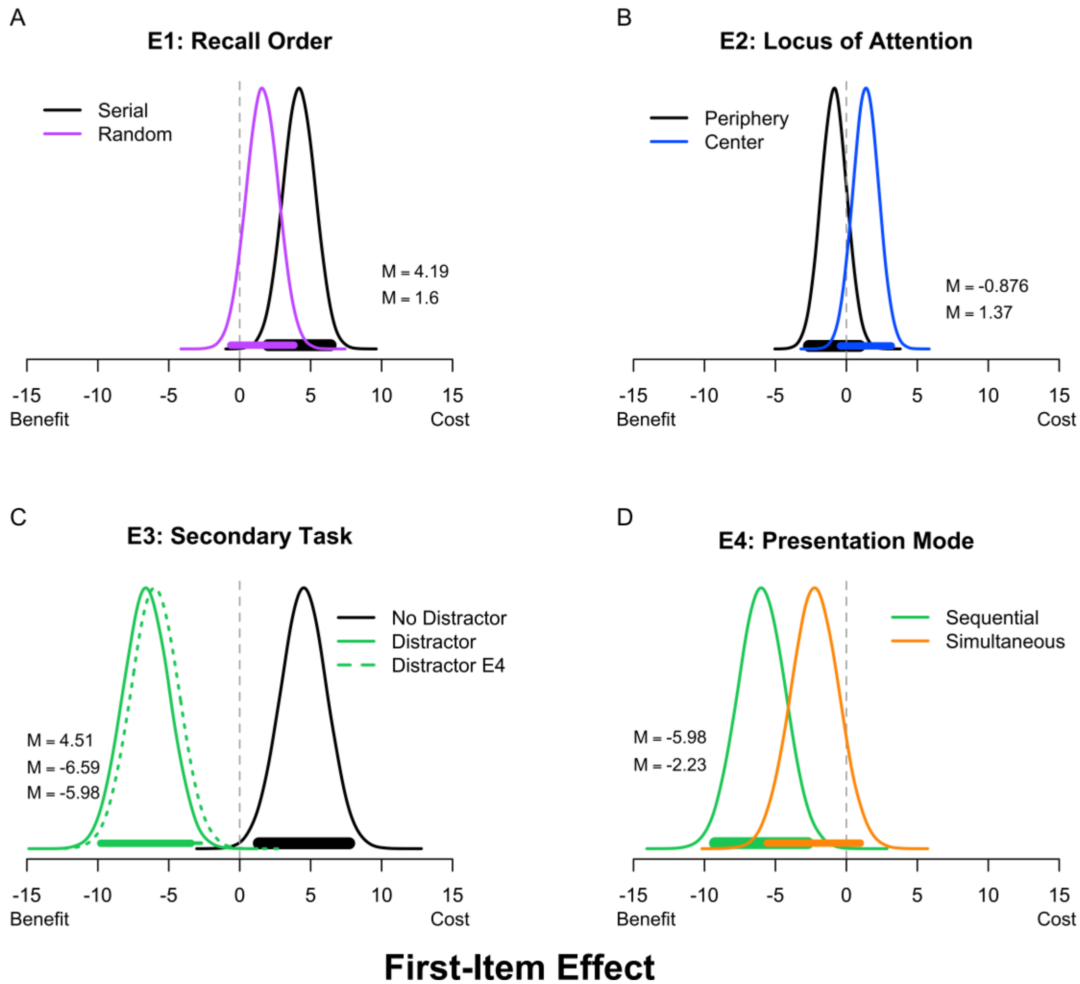
Error in Recalling the First-Item in the Short and Long Consolidation Intervals, and Evidence for the Consolidation-Time Effect

		Recall Error			Consolidation Effect
Exp.	**Condition Name**	**Short**	**Long**	**Δ**	**Evidence (BF_{10}) Long < Short**
E1	Serial Recall	32.52 (3.32)	36.71 (2.76)	+4.19	0.04
	Random Recall	55.63 (2.35)	57.23 (2.60)	+1.60	0.08
E2	Periphery	25.74 (3.68)	24.86 (2.54)	+0.88	0.24
	Center	14.03 (2.54)	15.41 (2.52)	+1.37	0.06
E3	No-distractor	24.00 (3.86)	28.54 (3.49)	+4.54	0.06
	Distractor	51.42 (3.85)	44.80 (3.68)	-6.62	132.44
E4	Distractor + Sequential	46.87 (3.52)	40.86 (4.02)	-6.01	20.84
	Distractor + Simultaneous	33.02 (4.86)	30.79 (4.96)	-2.23	1.43

Note. Values below 0.33 indicate substantial evidence for the Null. Values above 3 indicate substantial evidence for the alternative hypothesis.

Figure 3

Posterior Distribution of the Consolidation Time Effect for all Experiments



Note. Consolidation time effect = difference in recall error between the short and long consolidation conditions for the first item (e.g., first-item effect). The bar underneath the curve represent the 95% HDI of the distribution. M = mean of the posterior.

8.3.5.2. Analysis Over all Input Positions

In the previous section, we focused on the effect of consolidation time on the first-item. The assumption that consolidation follows a ballistic process, however, is also obtained from the observation that the first item does not show a consolidation time effect, whereas the remaining ones benefit from consolidation. Accordingly, Ricker and Hardman (2017) presented evidence for a two-way interaction between consolidation time and serial input position. This pattern is visualized in Figures 2A and 2B for the serial and the random recall tests, respectively.

To test this, we conducted two BANOVAs (one for each recall test) having consolidation time and input position as predictors (see Table 4). For both recall tests, the best model was the full model including all predictors, and there was overwhelming evidence to retain all predictors in the model. The main effect of consolidation time reflects the findings that short consolidation resulted in worse performance than long consolidation – a consolidation time benefit. The main effect of input position reflects the observation of better performance for the first item and gradual increase in recall error for the remaining items. The two-way interaction reflects the finding that the consolidation time benefit differs over serial input positions, and this was mainly driven by the first item, as our previous analysis already revealed.

Finally, we ran a 3-way BANOVA to investigate whether the three-way interaction was not included in the best model when considering the data of both recall tests together. The BANOVA included the three fixed factors of input position (levels: 1-4), consolidation time (levels: 200 ms, 2000 ms) and recall test (levels: serial order, random order). The best model included all three main effects and two-way interactions of memory test × input position and consolidation time × input position ($BF_{10} = 1.09 \times 10^{101}$). Critically, the best model was favored with a $BF_{10} = 9.01$ against the full model including the three-way interaction, indicating that consolidation time changed similarly across input position for both recall tests.

8.3.5.3. Analysis on Output Position for Random Recall

Figure 2C shows recall across the four output positions in the random recall test. Research shows that memory recall is negatively impacted by testing: as the output position increases, so does the error rate, a phenomenon known as output interference. Here we sought to investigate whether protection from output interference contributes to the first-item effect. In the serial recall test, the first-presented item is always tested first, and hence it is always protected from output interference. In contrast, the subsequently tested items would suffer from output interference of the previously tested items.

In our preregistration, we mentioned that we would run this analysis if there was no first-item effect in the random recall test. However, we decided to still run this analysis to explore whether there was some evidence of a differential effect of consolidation time over output position.

We ran a BANOVA with the two fixed predictors of consolidation time and output position, to test whether the interaction would be included in the model or not (see Table 4). To specify, the inclusion of an interaction would mainly be driven by the absence of a consolidation time benefit on the first output position - and the inclusion of an interaction would therefore strengthen this finding. The best model was again the full model. However, the inclusion of the interaction was only favored by a $BF_{10} = 4.52$, which provides modest evidence for an interaction. A Bayesian *t*-test on the first output position revealed ambiguous evidence *against* a consolidation time benefit ($BF_{10} = 0.35$; $BF_{01} = 2.86$). This result suggests that part of the consolidation-time benefit may be due to protection from output interference.

Table 4

Bayes Factor (BF) of Models with Different Fixed Effects Over the Null, and BF of the Best Model Over the Model Specified in Each Row (BF_{Best}/BF_{Mrow}) for the Recall Error Data over Input and Output Positions in Experiment 1

Memory Test + Position	Model n°	Included Fixed Effects: Consolidation	Position	Cons. x Position	BF_{10}	BF_{Best}/BF_{Mrow}
Serial – Input Position	**1**	✓	✓	✓	$\mathbf{2.73 \times 10^{72}}$	**1**
	2	✓	✓	---	7.32×10^{62}	3.72×10^{9}
	3	✓	---	---	1.64×10^{4}	1.66×10^{68}
	4	---	✓	---	3.27×10^{52}	8.33×10^{19}
Random – Input Position	**1**	✓	✓	✓	$\mathbf{6.44 \times 10^{15}}$	**1**
	2	✓	✓	---	2.74×10^{13}	2.35×10^{2}
	3	✓	---	---	3.59×10^{4}	1.83×10^{11}
	4	---	✓	---	1.09×10^{8}	$5.89 \text{ x } 10^{7}$
Random – Output Position	**1**	✓	✓	✓	$\mathbf{3.92 \times 10^{45}}$	**1**
	2	✓	✓	---	8.69×10^{44}	4.52
	3	✓	---	---	1.88×10^{4}	2.08×10^{41}
	4	---	✓	---	5.95×10^{36}	$6.60 \text{ x } 10^{8}$

Note. ✓ = effect included in the model. Best model is depicted in bod font. Best model = models with higher BF over the Null.

8.3.5.4. Categorical vs. Continuous Mixture Model

As mentioned in the preregistration, we also modeled the responses using the Bayesian hierarchical categorical-continuous mixture model (Hardman et al., 2017) with the CatCont package (Hardman, 2016) implemented in R. The model assumes that responses are either informed by memory (P^M) or reflect guessing (1- P^M). Responses informed by memory could reflect continuous (P^O) or categorical (1 - P^O) information about the item. Continuous information allows for a fine-grained response that varies linearly with the studied feature. The continuous response can be more or less fine-grained – which reflects the continuous imprecision (σ^O). In contrast, categorical responses cluster around some canonical values (e.g. right, left) along the feature space. The model assumes two sources of guessing: categorical guessing, when participants randomly guess prototypical angles (P^{AG}), or continuous guessing, indicated by a uniform distribution along the feature space (1 - P^{AG}). We wanted to assess how consolidation time changed the probability of responses informed by categorical as opposed to continuous information as well as the continuous imprecision of the memory representation across serial and random recall. To get continuous information, P^M needed to be multiplied by P^O, whereas for categorical information it is: P^M x (1-P^O). The continuous imprecision parameter (σ^O) was used as outputted by the model.

We fitted a the between-item variant model of the package that allowed the three main parameters P^M, P^O, and σ^O in the model to vary across experimental conditions. The parameter values and distributional probabilities were determined through Markov chain Monte Carlo (MCMC) sampling techniques. The model was run with 10,000 of these iterations, of which 1,000 were disregarded as burn-in iterations and the model fit the data quite well (for a posterior predictive check, see Supplementary Materials at https://osf.io/gncbq/).

In short, more consolidation time was beneficial for both, categorical and continuous memory but it generally did not influence continuous imprecision for both the serial and random recall conditions, in line with the results reported by Ricker and Hardman (2017). Given that effects on mixture model parameters were not critical to address our main question, we present the results of this analysis in the Supplementary Materials available on the OSF of this project (https://osf.io/gncbq/). Furthermore, we did not proceed into following up on the mixture modeling in the remaining experiments as our main aim of this study was to assess whether consolidation was a ballistic process or not.

8.3.6. Discussion

In Experiment 1, we replicated the study Ricker and Hardman (2017) with two additions. First, we varied whether recall was completed in serial or random order to examine whether the first-item effect could arise from strategies for recall preparation and protection from output interference. Second, we included an articulatory suppression procedure to rule out that consolidation time effects arise from verbal labeling.

To sum up, we replicated the first-item effect found in Ricker and Hardman (2017) for forward serial recall order even under a suppression procedure. Second, we found that this effect did not disappear with a random recall test. Hence, we found no evidence that the pattern of all-or-none consolidation for the first-presented item was due to strategic preparation for the serial recall test.

In an explorative analysis, we found some evidence that output interference may partly contribute to the first-item effect found in the random recall test. This is because when we analyzed recall error as a function of output position only, we observed that items recalled first did not show a consolidation time effect. As explained previously, the forward serial recall procedure entails a confound between input and output position, such that the first item tested is immune to output interference. Consolidation serving to protection from output interference effects would thus have predicted the same pattern of results as obtained for the first-item effect. The findings of the random recall test in Experiment 1 suggests that the first-item effect in serial recall may also partially be explained by a lack of output position interference for this item.

8.4. Experiment 2

Experiment 2 aimed to test one alternative explanation for the lack of a consolidation effect on the first-item, namely that this is related to the requirements to shift visuospatial attention to search for the location where the next item was presented. Ricker and Hardman (2017) and our Experiment 1 presented items in a random subset of 8 locations. Hence it is possible that the first item was always consolidated in these studies because participants did not shift their visuospatial attention to the location of the next item. Two features of task set-up could have contributed to the lack of motivation to quickly shift attention to the subsequent item: (a) participants did not know where the next item would be presented, or (b) visuospatial attention was slow to disengage from the first item and to re-engage to the next item. If all information would have been presented in predictable locations or at the same location, consolidation of the first item might be interrupted by the presentation of the subsequent item.

Accordingly, in Experiment 2, we examined whether predictability of the spatial location of the next item or the requirement to shift visuospatial attention contributes to the first-item effect. To assess this, we varied whether items were presented in different but predictable spatial locations (hereafter the periphery condition) or all in same location, namely the screen center (center condition).

This experiment was preregistered under https://osf.io/dgnxz/. Data, materials and analysis scripts are openly accessible at https://osf.io/gncbq/.

8.4.1. Method

8.4.1.1. Participants

Experiments 2 to 4 were conducted as part of the requirements to earn credits on an Introductory Course on Research Methods in Experimental Psychology in which undergraduate (Bachelor) students learned how to plan an experiment, pre-register it, collect the data, analyze it, and report it in the form of posters and a research paper. The sample sizes across Experiments 2 to 4 were defined by the number of undergraduate students assigned to each experiment: each student needed to collect data of 12 participants, and 3 to 4 students were assigned to each experiment, resulting in a sample size of either 36 or 48 participants. The sample in Experiment 2 comprised 48 students (M = 23.92 years old; SD = 4.01; 17 women). All participants were naïve to the hypotheses tested in the study, and they were not taking part on the research methods course.

8.4.1.2. Materials and Procedure

The design of Experiments 2 to 4 was similar to Experiment 1, with the following changes. First, item locations were reduced from eight to four. We selected the four canonical values: top, right, bottom and left. The items were presented in yellow (RGB = 255 255 0), pink (RGB = 255 50 255), blue (RGB = 0 0 255), and red (RGB = 255 0 0). In addition, the four items were always presented from the top onward in clockwise order. This was also the order of testing in all following experiments. Thus, the items were from now on always tested in forward serial order.

The mask remained the same as in Experiment 1. The same feedback screen as in Experiment 1 was presented for 2 seconds. This feedback also included the mean recall error across the four memory items which was presented in the middle of the screen.

Lastly, participants were tested in a group-setting, meaning that up to four participants could be concurrently tested. Participants were instructed to remain quiet throughout the whole experimental procedure.

Experiment 2 included two presentation mode conditions: The memory items and mask were either presented in the four locations surrounding the center of the screen (periphery condition) similarly to Experiment 1, or they were presented in the middle of the screen (center condition, see Figure 1C). The presentation mode conditions were organized into two blocks, each containing 72 test and 8 practice trials. The order of blocks was counterbalanced across participants.

8.4.1.3. Preregistered and Not Preregistered Analyses

In the preregistrations of Experiments 2 to 4, we stated that we would use Null-hypothesis inferential statistics to assess performance. This was due to these preregistrations being conducted by the undergraduate students which were only familiar with these set of statistical procedures. The results of these analyses are available under the Supplementary Analysis at https://osf.io/gncbq/. Here, we used Bayesian inferential statistics to assess the evidence for the pre-registered hypotheses. Both analyses were generally in agreement.

For Experiment 2, a 3-way ANOVA (i.e., serial input position x consolidation time x presentation mode) was preregistered, which was selected by the undergraduate students. As this was an undergraduate course, it encouraged students to think and develop their own analysis pattern. The goal of this analysis was to test whether the novel manipulation of presentation mode introduced in the present study moderated the interaction of serial input position and consolidation time observed by Ricker and Hardman (2017). We will present the results of a 3-way BANOVA towards the end of our results section.

In this paper, however, we decided that it aided understanding to present a more focused test of the first-item effect first. Therefore, our results section starts with a 2-way BANOVA on the data of the first input position testing the effect of consolidation time x presentation mode, followed by two separate Bayesian *t*-tests contrasting the effect of consolidation time separately for each presentation mode. Next, we report two separate 2-way BANOVAs which tested the interaction of serial input position and consolidation time in each presentation mode condition on its own. Finally, the preregistered 3-way ANOVA is presented (here BANOVA).

8.4.2. Results

Figure 4 presents recall error in each serial input position as a function of consolidation time and presentation mode. As in Experiment 1, we will first evaluate the evidence for a consolidation time effect in the first presented item, and then considering all serial positions.

8.4.2.1. First-item effect

Figure 4 shows no evidence of an improvement in performance as a function of longer consolidation time for the first item irrespective of presentation mode. As in the previous experiment, we ran a BANOVA on this data. The best model included only the main effect of presentation mode ($BF_{10} = 9.58 \times 10^{54}$). This model was favored against the model including the interaction ($BF_{10} = 202.02$), both main effects ($BF_{10} = 34.55$) and consolidation time only ($BF_{10} = 3.21 \times 10^{56}$). The effect of presentation mode reflects the observation of better performance in the center condition than in the periphery condition. The exclusion of consolidation time from the best model is in line with a first-item effect.

The posteriors of the first-item effect for each presentation condition are presented in Figure 3B. For both presentation modes, there was no evidence of a credible benefit or cost with longer consolidation time, and the HDIs of the conditions overlap substantially.

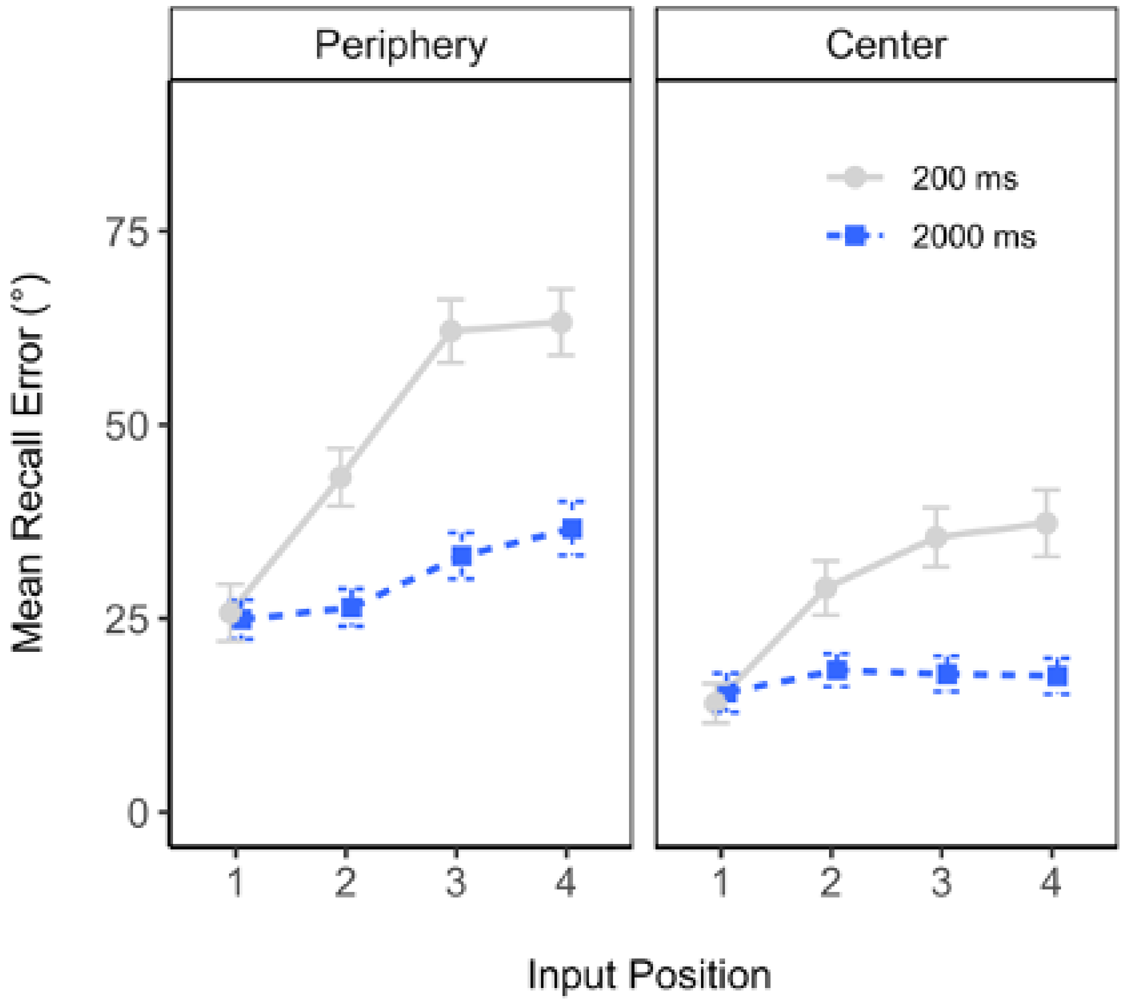
To further assess the evidence for the first-item effect we conducted two separate one-sided Bayesian *t*-tests assessing evidence for a consolidation time benefit on the first-presented item (see Table 3). For both presentation conditions, we found evidence against a benefit of consolidation time.

8.4.2.2. Analysis over All Input Positions

Figure 4 shows that consolidation time improved recall across items 2-4, but not the first item in both presentation modes. We conducted two separate BANOVAs on the data of each presentation mode, where consolidation time (200 ms vs. 2000 ms) and input position (levels: 1-4) were set as fixed predictors. Results are presented in Table 5.

Figure 4

Recall Error over Input Position and Consolidation Time in Experiment 2



Note. The error bars represent the 95% within-subjects confidence interval.

Replicating Experiment 1, the best model in both conditions included the main effects and the interaction between serial input position and consolidation time.

Table 5

Bayes Factor (BF) of Models with Different Fixed Effects Over the Null, and BF of the Best Model Over the Model Specified in Each Row (BF_{Best}/BF_{Mrow}) for the Recall Error Data over Input Position in Experiment 2

		Included Fixed Effects				
Condition	Model n°	Consolidation	Position	Cons. x Position	BF_{10}	BF_{Best}/BF_{Mrow}
Periphery	**1**	✓	✓	✓	$\mathbf{6.18 \times 10^{66}}$	**1**
	2	✓	✓	---	2.79×10^{53}	2.21×10^{13}
	3	✓	---	---	1.03×10^{20}	6.01×10^{46}
	4	---	✓	---	4.30×10^{22}	1.44×10^{44}
Center	**1**	✓	✓	✓	$\mathbf{7.17 \times 10^{37}}$	**1**
	2	✓	✓	---	1.51×10^{28}	4.76×10^{9}
	3	✓	---	---	3.19×10^{15}	2.25×10^{22}
	4	---	✓	---	2.20×10^{9}	3.26×10^{28}

Note. ✓ = effect included in the model. The best model is printed in bold font. Best model = model with highest BF over the Null.

Lastly, we ran a 3-way BANOVA, including the fixed predictors of consolidation time, input position, and presentation mode. The best model included all main effects and all two-way interactions ($BF_{10} = 4.88 \times 10^{147}$). It was favored against the second-best model that in addition included the three-way interaction by a BF_{10} of 8.75 [1].

[1] In the Frequentist analysis this 3-way interaction was $p = 0.018$

8.4.3. Discussion

In Experiment 2, we tested the hypothesis that consolidation of the first presented item was not interrupted by the subsequent item because the latter appeared in a different and unpredictable spatial location, and participants did not quickly disengage their attention from item 1 to move to item 2. If this was the case, presentation of the subsequent memory item at predictable locations or at the same location as the first one (thereby not requiring shifts of visuospatial attention) should stop ongoing consolidation. To test for this, the memory items were either presented in predictable locations in the periphery, or in the middle of the screen at a fixed location.

First, we replicated the first-item effect found in Experiment 1 in the periphery conditions, thereby showing the consistency of this effect across slightly different experimental set-ups (e.g., random set of locations vs. fixed set of locations). Second, we found evidence for a first-item effect in the central presentation mode condition, where the items were always presented in the middle of the screen. This indicates that neither uncertainty in the location of the items nor the requirement to shift visuospatial attention across the screen where conditions that explained the lack of a consolidation effect for the first item.

8.4.3.1. Grouping effect in Experiment 2

In Experiment 2, we observed that memory performance for the central presentation condition was clearly better than for memory items presented in different spatial locations. Previous studies have suggested that presentation of all items at the same location would imply a cost for working memory performance (Makovski, 2016). This is in line with the idea that increasing the number of retrieval cues would improve performance in visual working memory tasks (Bae & Flombaum, 2013). Here we observed the opposite: the same location condition (center) was better than the different locations condition (periphery), even though the different location condition increased the number of retrieval cues along with the spatial separability of the items.

One possible explanation for this unexpected finding is that participants in the center condition could focus their attention only on the dots rather than the combination of the dot and ring. In this case, participants could have grouped the four dot locations on one single ring. It has been shown that compression (Brady et al., 2009) or chunking (Thalmann et al., 2019; see also, Huang & Awh, 2018) of memory items into one memory representation is beneficial for working memory. Here, participants may have used the movements of the dots across a single reference frame to chunk the items together thereby benefiting performance.

8.5. Experiment 3

Experiment 3 tested the hypothesis that a secondary task could interrupt the consolidation of a previously presented item. We reasoned that the first-item effect might be due to the presentation of the second item not being sufficiently demanding to withdraw attention from the first item. Studies using a secondary task, in contrast, require an immediate response to the secondary task, which in turn might force participants to withdraw attention from consolidation. Nieuwenstein and Wyble (2014) presented evidence consistent with this possibility: they observed that a distractor task interfered with memory performance, and more so, when the secondary task occurred shortly after the memory array (see also, Bayliss et al., 2015; Kandemir et al., 2017). The set-up of this study was, however, very different from the one of Ricker and Hardman (2017), and it is still unclear whether the secondary task is indeed responsible for their opposite findings regarding consolidation. Accordingly, the aim of Experiment 3 was to assess for the role of a distractor task in stopping consolidation using the same design as in Ricker and Hardman (2017) and our previous experiments.

Experiment 3 was preregistered at https://osf.io/6s5fr/. Data, materials and analysis scripts are openly accessible at https://osf.io/gncbq/.

8.5.1. Method

8.5.1.1. Participants

Thirty-six students (*M* = 22.53 years old; *SD* = 2.79; 36 women) participated in Experiment 3. Data of three participants was excluded from the main analysis because they failed to correctly perform the distractor task: One participant did not follow the instructions of the distractor task and the other two did not reach our threshold of 90% correct responses overall on the distractor task (final *N* = 33).

8.5.1.2. Materials and Procedure

In this experiment, participants first trained the distractor task for 32 trials. The distractor task consisted of a parity judgment task, in which a visually presented digit (1-4, 6-9) was classified as odd or even by a left or right mouse click, respectively. Participants were given a maximum of 5000 ms to respond to each digit. If they failed to do so, a time-out was recorded and the program moved on. After every response, a feedback was shown for 150 ms, stating either “correct” in green or “wrong” in red both in German.

This experiment contained two blocks: a no-distractor block and a distractor block. The no-distractor block was identical to the periphery condition in Experiment 2.

In the distractor block, the above described distractor task was inserted immediately after the consolidation interval of each item (see Figure 1D). Participants needed to respond within 5000 ms whether the digit presented at the location of the last encoded item was odd or even. If participants failed to respond within 5000 ms, a time-out was recorded and the next item was presented. In case a response was registered, the next item was presented.

As in Experiment 2, at the end of the trial feedback was displayed for 2 s. The order of the no distractor and distractor blocks were counterbalanced across participants. As in Experiment 2, each block contained 72 test trials and 8 practice trials.

8.5.1.3. Preregistered and Not Preregistered Analyses

Similarly to Experiment 2, the preregistered analysis was a 3-way ANOVA, but including only the first two input positions for this factor. We added all input positions to estimate the full effect of the interaction across all input positions. Furthermore, we analyzed the data with the same approach as stated under the same section in Experiment 2, which were not preregistered: the BANOVA on the first input position testing the effect of consolidation time x distractor condition, then two separate Bayesian *t*-tests on the first input position. Then we followed with two BANOVAs testing the interaction of input position x consolidation time in each distractor condition separately. Finally the 3-way ANOVA is presented.

8.5.2. Results

8.5.2.1. Distractor Task

Figure 5B shows that participants performed the distractor task at a high accuracy rate ($M = 0.96$, $SD = 0.01$). A BANOVA having consolidation time and input positions as predictors revealed that the best model included only a main effect of consolidation (BF_{10} = 3.74), but this model only received modest support. This model was favored by the model including all main effects and the interaction (BF_{10} = 7.64), only the main effects (BF_{10} = 5.32) and only the effect of input position (BF_{10} = 20.98). A Bayesian *t*-test revealed that this difference in consolidation time seems to primarily emerge from the first input position (BF_{10} = 12.12), and not input positions 2 (BF_{10} = 0.21), 3 (BF_{10} = 0.33), and 4 (BF_{10} = 0.36).

Figure 5C presents reaction times (RT) to the distractor task in each serial input position. A BANOVA having consolidation time and input position as predictors favored the model including only the main effect consolidation time ($BF_{10} = 1.92 \times 10^{17}$). This model was favored against the model including both main effects and the interaction term (BF_{10} = 35.78), the model including only the two main effects (BF_{10} = 2.56) and the input position only model ($BF_{10} = 1.31 \times 10^{18}$). Hence, RTs were overall higher in the short consolidation time condition (M = 1.01, SD = 0.04) than in the long consolidation time condition (M = 0.78, SD = 0.03). But it is unclear, whether distractor task RTs further changed across the four input positions.

8.5.2.2. First-Item Effect

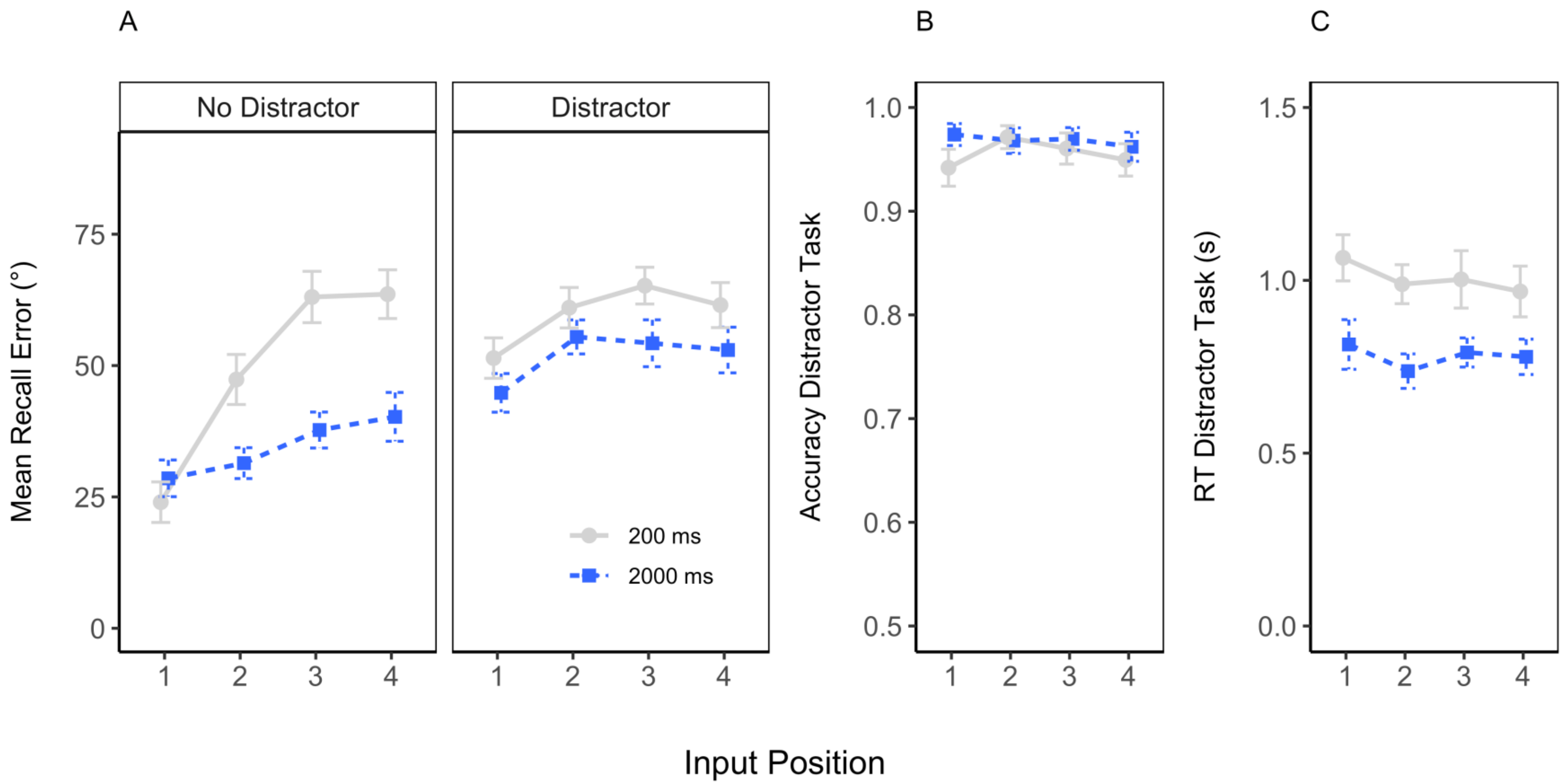
Figure 5A presents serial position curves for the conditions with and without a distractor task. Here we will focus on the first serial position. A BANOVA having consolidation time and the distractor task (levels: no distractor, distractor) as predictors, indicated that the best model included the two main effects and their interaction (BF_{10} =

2.28×10^{75}) [2]. This model was favored against the model including the two main effects ($BF_{10} = 5287.04$) and the two models including either only a main effect of distractor task condition ($BF_{10} = 257.99$) or consolidation ($BF_{10} = 4.86 \times 10^{76}$). It is important to mention here again that for the no distractor task condition, the difference in consolidation time seems to emerge into the opposing direction of what we would predict, as the performance in the long consolidation time interval seems to be worse than in the short consolidation time interval.

[2] In the Frequentist analysis the consolidation factor was not significant (p = 0.446)

Figure 5

Recall Error (Panel A) in the Memory Task, and Accuracy (Pane B) and Reaction Time (RT, Panel C) in the Distractor Task in Experiment 3



Note. The error bars represent the 95% within-subjects confidence interval.

The respective posteriors are presented in Figure 3E. For the no distraction condition, more consolidation time was associated with an increase in recall error (consolidation time cost), whereas for the distractor condition, more consolidation time yielded lower recall error (consolidation time benefit). For the first time, we can see that the HDIs of the two posteriors do not overlap. Moreover, we can see that the HDIs do not include zero.

Table 3 presents the evidence for a consolidation time effect in each condition (no distractor vs. distractor). For the no-distractor condition, we replicated the lack of a consolidation time benefit in the first serial input position. In contrast, for the distractor condition, there was overwhelming evidence for a consolidation time benefit on the first input position. Hence, Experiment 4 shows that consolidation of the first item is interrupted when a distractor task appears shortly after this item was presented. This is critical evidence against the assumption that consolidation is a ballistic process.

8.5.2.3. Analysis over All Input Positions

We then ran two BANOVAs having consolidation time and input position as predictors separately on the data of each distractor condition (see Table 6). The best model for the no distractor condition included both main effects and their interaction, and this model was clearly favored over the model containing only the main effects. In contrast, the best model of the distractor condition included only the main effects of consolidation and serial position, but there was substantial evidence against the inclusion of the interaction term (see Table 6).

Finally, we ran a 3-factorial BANOVA, including the fixed factors consolidation time, input position, and distractor condition. The best model clearly favored all main effects and all possible interactions, e.g., the full model ($BF_{10} = 2.39 \times 10^{78}$), and this model was favored over the model excluding the three-way interaction with a $BF_{10} = 5.17 \times 10^{3}$. This provides evidence supporting the assumption that the distractor task condition equated all serial position in terms of the consolidation time effect.

Table 6

Bayes Factor (BF) of Models with Different Fixed Effects Over the Null, and BF of the Best Model Over the Model Specified in Each Row (BF_{Best}/BF_{Mrow}) for Recall Data over Input Position in Experiment 3

		Included Fixed Effects				
Condition	Model n°	Consolidation	Position	Cons. x Position	BF_{10}	BF_{Best}/BF_{Mrow}
No Distractor	**1**	✓	✓	✓	$\mathbf{6.79 \times 10^{45}}$	**1**
	2	✓	✓	---	2.61×10^{34}	2.60×10^{11}
	3	✓	---	---	2.15×10^{9}	3.16×10^{36}
	4	---	✓	---	4.65×10^{18}	1.46×10^{27}
Distractor	1	✓	✓	✓	1.32×10^{12}	8.99
	2	✓	✓	**---**	$\mathbf{1.18 \times 10^{13}}$	**1**
	3	✓	---	---	3.48×10^{5}	3.40×10^{7}
	4	---	✓	---	1.77×10^{6}	6.69×10^{6}

Note. ✓ = effect included in the model. The best model is printed in bold font. Best model = model with highest BF over the Null.

8.5.3. Discussion

In Experiment 3, we tested whether the first-item effect would disappear when a distractor task was inserted shortly after each memory item because this attentionally demanding task draws attention away from the memory trace thereby interrupting ongoing consolidation. To test for this, participants were exposed to two conditions: a condition without a distractor task (replicating Experiment 2) and a condition in which a distractor task was inserted after each memory item.

Results of the distractor condition showed, for the first time, a consolidation time effect across all four serial input positions. The absence of the first-item effect in the distractor condition indicates that consolidation of the first item was interrupted by the processing of the parity judgement task. When the distractor task was removed in the no-distractor task condition, we observed again the first-item effect. This finding challenges the assumption that consolidation is a ballistic process – consolidation of a memory item could be interrupted if a secondary, attentionally demanding task is presented shortly after encoding.

Furthermore, the distractor task was also impaired by the concurrent memory task: accuracy was lower (especially on the first input position) and RTs were slower under the short consolidation time. This shows that the memory task and the distractor task both competed with each other, yielding costs for both tasks when they were processed concurrently.

8.6. Experiment 4

Experiment 3 showed for the first time that consolidation can be interrupted provided that a distractor task is presented shortly after each memory item. This suggests that the results of Nieuwenstein and Wyble (2014) indeed reflect the interruption of consolidation by a trailing two-choice RT task. As we pointed out before, their experiment not only involved the presentation of a distractor task, but also the simultaneous presentation of memory items in a verbal task involving letters or Chinese symbols. All of these features make it unclear whether their results could be generalizable to a sequential visuospatial task as the one used by Ricker and Hardman (2017).

The general goal of Experiment 4 was to replicate the distractor condition of Experiment 3 which used a sequential display, and assess whether the same results would hold when the memoranda were presented simultaneously, similarly to the procedure of Nieuwenstein and Wyble (2014). Furthermore, we wanted to address one potential confound on the analyses reported by Nieuwenstein and Wyble (2014). They averaged recall performance across all output positions in their study, although items were recalled in a strict order (e.g., from left to right). It is thus unclear whether they would have found evidence for a first-item effect, as participants correctly reported, on average, 2.5 letters (out of 4) for the short consolidation interval. This could mean that participants always reported the first outputted item correctly irrespectively of consolidation time. In the present study, we aimed to assess whether the effect of consolidation time with a simultaneous array is evident for all output positions.

This experiment was preregistered under https://osf.io/43zvq/. Data, materials and analysis scripts are openly accessible at https://osf.io/gncbq/.

8.6.1. Method

8.6.1.1. Participants

Thirty-six students (*M* = 22.70 years old; *SD* = 2.76; 22 women) took part in Experiment 4. Two participants were excluded from the dataset: one participant due to program crash and the other due to failure to reach the threshold of 90% correct responses overall on the distractor task, leaving a final sample of *N* = 34.

8.6.1.2. Materials and Procedure

This experiment again consisted of two blocks that were counterbalanced across participants. The distractor-sequential block was identical to the distractor task condition of Experiment 3. The distractor-simultaneous block worked as follows (see Figure 1E): the four memory items were simultaneously presented for 600 ms (e.g., 150 ms per item). All four items were then masked for 50 ms. After the consolidation time interval of either 200 or 2000 ms, the same distractor digit was presented in all four item locations simultaneously. Participants had a maximum of 5000 ms to classify this digit as odd or even and press the corresponding mouse button. The response (or the time-out) was followed by the presentation of the test display. At test, items were always recalled in the same order, namely from the top position (yellow item in Figure 1E) to the left position (red item) in clockwise order. Each block consisted of 72 test trials and 8 practice trials. Before the start of the experimental blocks, the distractor task was first practiced as done in Experiment 3.

8.6.1.3. Preregistered and Not Preregistered Analyses

For the same reasons as mention in the previous Experiments 2 and 3, only a 3-way ANOVA was preregistered. We analyzed the data using the same gradual approach as stated under the same section in Experiment 2, which were not preregistered: the BANOVA on the first input position, followed up with Bayesian *t*-tests, two BANOVAs assessing the interaction of serial input position and consolidation time, and lastly the 3-way BANOVA.

8.6.2. Results

8.6.2.1. Distractor Task

Figure 6B presents distractor task accuracy over serial positions in the distractor sequential condition, and for the simultaneous condition (All category). Accuracy in the distractor task was generally high ($M = 0.96$, $SD = 0.02$). The respective distractor task RTs are presented in Figure 6C. Overall RTs, e.g., considering both the sequential and simultaneous presentation mode, were slower in the short consolidation time interval ($M = 1.04$, $SD = 0.04$) than in the long consolidation time interval ($M = 0.76$, $SD = 0.02$).

We first computed a BANOVA on distractor task accuracy in the sequential condition. The best model included the main effects of consolidation time and input/output position ($BF_{10} = 190.52$). However, there was ambiguous evidence whether the factor input position needed to be included in the model ($BF_{10} = 1.17$). This best model was further favored by the model including the interaction ($BF_{10} = 14.70$) and the model including only input position ($BF_{10} = 208.67$). The model on the RT data for the sequential task condition clearly favored inclusion only of the main effect of consolidation ($BF_{10} = 1.18\times 10^{28}$) over the model including all terms ($BF_{10} = 465.10$), the two main effects ($BF_{10} = 28.37$), and only input position ($BF_{10} = 4.90 \times 10^{29}$). Hence, participants were less accurate and took longer to respond within the short consolidation time interval than in the short interval.

For the simultaneous condition, Bayesian *t*-tests on distractor task accuracy (BF_{10} = 696.83) and reaction time ($BF_{10} = 3.61 \times 10^{21}$) revealed clear consolidation time effects. Hence, distractor task accuracy was worse and it took participants longer to respond in the short consolidation than in the long consolidation condition.

8.6.2.2. First-Item effect

Figure 6A presents serial position curves. Table 3 presents evidence for our directional *t*-tests on the consolidation time benefit. For the sequential condition, there was strong evidence in favor of a consolidation time benefit on the first item, replicating Experiment 3. The simultaneous condition yielded inconclusive evidence for an effect of consolidation time.

Finally, we ran a BANOVA having consolidation time and the two presentation modes as predictors. The best model included the two main effects ($BF_{10} = 2.33 \times 10^{21}$) and was favored against the model including also the interaction (BF_{10} = 6.26), the model including only the presentation condition (BF_{10} = 12.88) and the model including only consolidation time ($BF_{10} = 2.04 \times 10^{20}$). Thus, participants' performance was better in the sequential than the simultaneous condition for the first input position and it seems as if participants responded better in the long than the short consolidation interval. If we look at Figure 6B, however, we can see that for the simultaneous presentation mode there does not seem to be a difference between the two consolidation intervals on any input position.

To further investigate into this finding, Figure 3D shows the respective posteriors of the consolidation time effect. The posterior HDIs of both presentation mode conditions have a tendency to move towards a performance benefit, indicated by a lower recall error. This is especially visible for the sequential presentation condition, where zero is not included in the HDI, indicating that the short and long consolidation time interval differ, but this tendency is not fully credible for the simultaneous condition.

Relevant for our research question is the assessment of the sequential presentation condition in Experiment 4 in relation to the one in Experiment 3. In Figure 3C, we see that the sequential-distractor condition of Experiment 4 (dotted-green line) replicates the results of the distractor condition in Experiment 3 (solid-green line), yielding further evidence for the finding that a distractor task eliminated the first-item effect.

8.6.2.3. Analysis of all Input Positions

Table 7 presents the results of the BANOVA for each presentation mode condition separately. For the sequential presentation mode, the best model included the two main effects of consolidation and input position but not their interaction. This gives more credibility to the assumption that consolidation is not a ballistic process: it can be stopped with the use of an attentionally demanding task, like a parity judgement task.

For the simultaneous presentation mode, the best model only included output position. This is in line with our previous analysis not finding evidence for a consolidation time effect. The main effect of output position indicates that strong output interference with better performance for the first recalled item.

The last analysis, included a 3-way BANOVA of our fixed factors consolidation time (levels: 200 ms, 2000 ms), input/output position (levels: 1-4) and presentation mode (levels: sequential, simultaneous). The best model included all three main effects and two-way interactions of presentation condition x consolidation and presentation mode × input/output position, respectively ($BF_{10} = 3.70 \times 10^{67}$). But, the inclusion of the presentation condition × consolidation interaction into the model was favored only by a

BF_{10} = 1.06. The best model was favored over the full model including all possible terms by a BF_{10} = 389.44 [3]. The exclusion of the three-way interaction into the model suggests that consolidation time changed performance differently for the two presentation modes, in dependence of the output position.

Table 7

Bayes Factor (BF) of Models with Different Fixed Effects Over the Null, BF of the Best Model Over the Model Specified in Each Row (BF_{Best}/BF_{Mrow}) for the Recall Data over Input/Output Position Experiment 4

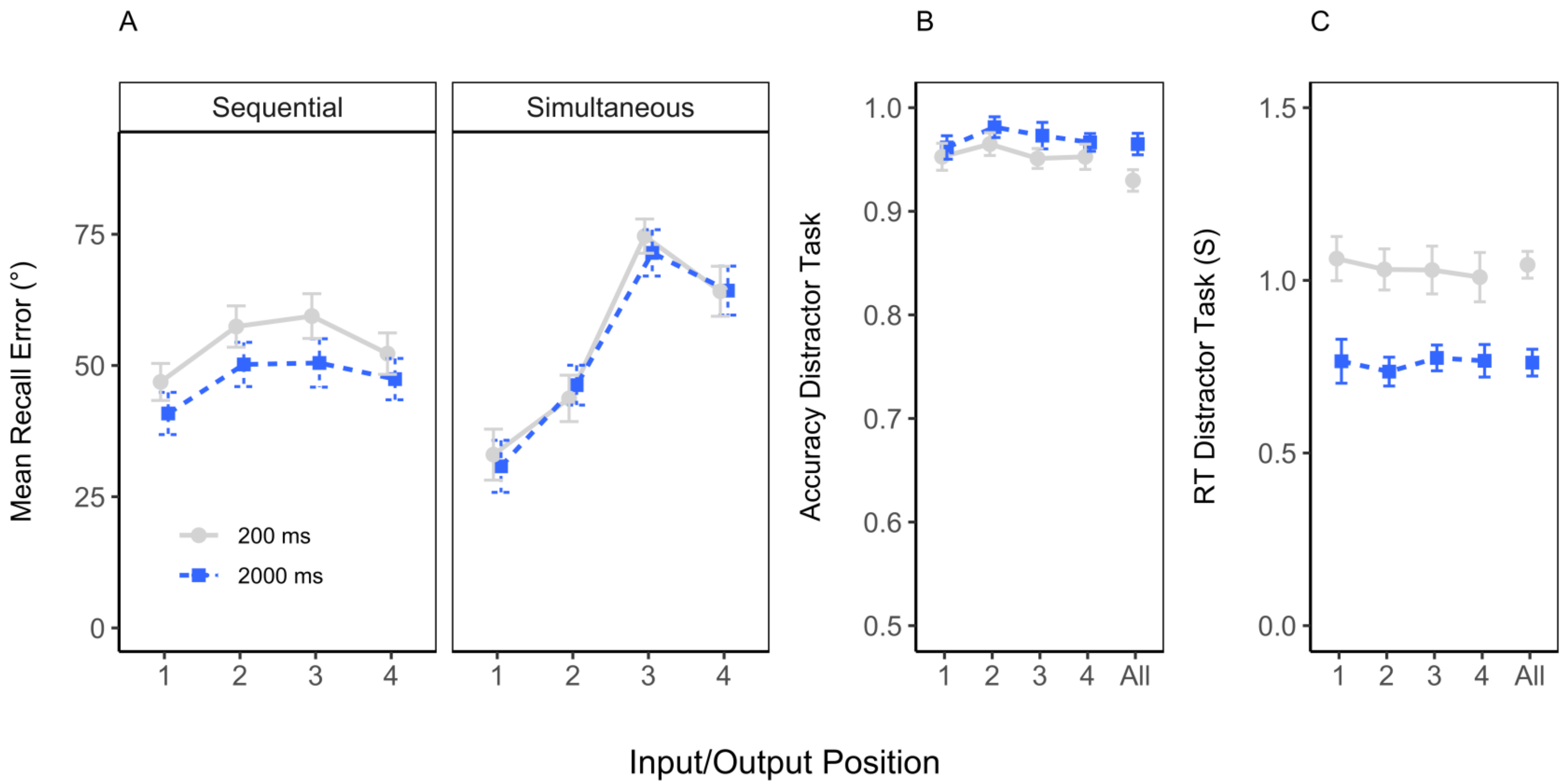
Condition	Model n°	Included Fixed Effects: Consolidation	Position	Cons. x Position	BF_{10}	BF_{Best}/BF_{Mrow}
Sequential	1	✓	✓	✓	1.71×10^{9}	14.73
	2	**✓**	**✓**	**---**	$\mathbf{2.52 \times 10^{10}}$	**1**
	3	✓	---	---	7.18×10^{3}	3.52×10^{6}
	4	---	✓	---	5.28×10^{5}	4.78×10^{4}
Simultaneous	1	✓	✓	✓	1.49×10^{54}	68.65
	2	✓	✓	---	1.53×10^{55}	6.70
	3	✓	---	---	0.14	7.40×10^{56}
	4	**---**	**✓**	**---**	$\mathbf{1.02 \times 10^{56}}$	**1**

Note. ✓ = effect included in the model. Best model is printed in bold. Best model = model with highest BF over the Null.

[3] The Frequentist analysis was not significant for the main effect presentation condition (p = 0.133), the interaction of input/output position × consolidation (p = 0.177) and as in the Bayesian ANOVA the three-way interaction (p = 0.385).

Figure 6

Recall Error (Panel A) in the Memory Task, and Accuracy (Panel B) and Reaction Time (RT, Panel C) in the Distractor Task in for Experiment 4



Note. The error bars represent the 95% within-subjects confidence interval. All = single-distractor processing episode following the simultaneous presentation of the memoranda.

8.6.3. Discussion

Experiment 4 replicated the distractor condition of Experiment 3 which involved sequential presentation of the memoranda followed by the distractor task. Furthermore, it tested whether consolidation is stopped when memory items are simultaneously presented, followed by a secondary distractor task. Our main goal was to assess whether we could replicate the results of Nieuwenstein and Wyble (2014) with a visuospatial task, and we aimed to more closely evaluate recall of each item in isolation. If participants encoded items in the order of testing even when they are simultaneously presented, a first-item effect might be evident even in the simultaneous condition.

The distractor-sequential condition of Experiment 4 replicated with a new sample the findings of Experiment 3 showing a consolidation time benefit across all four input positions and providing evidence against a first-item effect. These results further strengthen our finding that consolidation can be interrupted and show that consolidation is not a ballistic process.

With regards to the distractor-simultaneous condition, we did not find any evidence for a consolidation time effect for any item in any output position. We tested items always in clockwise order, starting from the top of the screen. We reasoned that such procedure could bias participants to encode items sequentially in clockwise order. We implemented such a procedure because we wondered whether the same bias would be present in the study of Nieuwenstein and Wyble (2014). However, we were not able to find a consolidation time effect across all four output positions.

8.6.3.1. Can Consolidation Occur in Parallel?

In Experiment 4, we presented the four memory items simultaneously for 600 ms to have equivalent total presentation duration between the simultaneous and the sequential condition in which each item was presented for 150 ms. Then, a mask was shown for 50 ms, followed by a consolidation interval of either 200 ms or 2000 ms. Accordingly, the shortest total duration separating item presentation and the onset of the distractor task was 850 ms. Across their series of experiments, Ricker and Hardman (2017) provided evidence that consolidation of one single memory item as used here (e.g., the orientation of a dot on the ring) asymptotes after 800 ms. Our original reasoning was that each item would be consolidated serially, even if presented simultaneously, following the logic of previous research that compared performance for simultaneous and sequential displays (Ricker & Cowan, 2014). These authors showed that equated total duration, as done here, tended to equate performance for simultaneous and sequential displays for the length of the retention interval used here. Their results and comparison between presentation mode conditions suggested that items were consolidated serially in visual working memory. Moreover, there are studies suggesting that orientation cannot be consolidated in parallel (Becker et al., 2013; Miller et al., 2014), supporting the idea that memory items would not have been consolidated within a short consolidation time interval of 600 ms. If items would have been consolidated serially in our experiment, full consolidation of all items would require 4 × 800 ms = 3200 ms. In this scenario, consolidation was unlikely to be fully completed either in the short or the long consolidation condition. However, if all items were processed in parallel, the 850 ms available in the short consolidation condition would already have provided more than enough time to fully consolidate all items, and we would not observe any effect of consolidation time. Hence, parallel consolidation can explain the lack of a consolidation time benefit across the two consolidation time intervals in the simultaneous condition (see also, Rideaux et al., 2018). Our choice to equate presentation duration over the two conditions therefore may have potentially created a confound that prevented us from observing a consolidation time effect in the simultaneous condition.

Parallel consolidation has been incorporated in Wyble et al. (2009, 2011), who presented a neural-network model: the *episodic simultaneous type-serial token model* (eSTST; see also, Bowman & Wyble, 2007). The model incorporates encoding of a memory item, its attentional selection, and working memory consolidation. This model assumes that incoming information activates the representation of a type (i.e., what the object is). Once the type reaches an activation threshold level, the type is bound to a token (i.e., a representation of the context in which the type occurred). The token itself is a working memory representation that sustains the activation of the type in order to make it retrievable later in working memory. This type-token binding-process is consolidation. The model further assumes that relevant information receives a transient attentional boost of activation of the type, thereby promoting its binding to a token, and hence its consolidation. The eSTST model assumes that multiple items can be bound in parallel if they are presented in close time proximity – because they can be selected simultaneously for consolidation. In Experiment 4, all four simultaneously presented memory items were relevant to the task as all of them needed to be reproduced, and the assumption that they were consolidated in parallel is congruent with the eSTST model.

Moreover, this model assumes that during consolidation, the allocation of attention to new information is suppressed, producing an attentional blink. When multiple items are consolidated simultaneously, they interfere with each other thereby also explaining capacity limitations.

Nieuwenstein and Wyble (2014) suggested that presentation of the distractor task was exciting enough to counteract the suppression produced by consolidation, thereby interrupting ongoing consolidation – and interfering with the memory task. This is in line with the results in our sequential presentation conditions in Experiments 3 and 4. If as the model predicts, simultaneously presented items were consolidated in parallel, this would explain why the distractor task did not disrupt consolidation in the simultaneous condition of Experiment 4.

In order to make clear conclusions on this point, however, future research needs to specifically test for the assumption of parallel consolidation. We deferred this question for future research because our aim in this study was to assess whether consolidation is a ballistic process and not whether multiple items can be consolidated in parallel.

8.7. General Discussion

With this series of experiments, we aimed to test whether ongoing consolidation of a visuospatial representation in working memory can be stopped once it has started. Consolidation was previously found to induce an attentional blink: consolidation of a first-presented memory item prevents encoding and consolidation of a subsequently memory item – a finding that we referred here to as the *first-item effect*. To explain this finding, Ricker and Hardman (2017) argued that consolidation in visual working memory is a ballistic process: once consolidation of an item started it could not be interrupted, leading to the blink of the second memory item.

In this series of experiments, we aimed to address the possibility that the first-item effect reflects a by-product of strategic preferences or features of the experimental design rather than a general property of the short-term consolidation process. We reasoned that features of the experimental set-up could have motivated participants to continue consolidation of the first item at the expense of the following item because: (a) they would be better prepared for the serial recall test, (b) they did not know where the next item would appear and shifting visuospatial attention around was costly, and (c) presentation of the second item does not demand an immediate response and hence it is not given proper priority. To address these possibilities, we ran four experiments in which we varied: (E1) whether recall was performed in serial or random order under verbal suppression; (E2) items were presented at predictable locations or the same location on the screen thereby reducing uncertainty and the need for shifts of visuospatial attention; and (E3-E4) whether a distractor task followed each memory item, or whether it followed the presentation of a simultaneous array.

The first-item effect was consistently observed across the serial and random recall order tests in Experiment 1, and across the predictable-peripheral locations and fixed center location conditions of Experiment 2. The results of these experiments supported the assumption that consolidation was inflexible and could not be interrupted. Only when a distractor task was imposed in-between presentation of the memory items in Experiments 3 and 4, we found clear evidence that ongoing consolidation was interrupted. This indicates that only imposition of a demanding task could motivate participants to withdraw attention from the first item, thereby interrupting its ongoing consolidation.

8.7.1. Interrupting Consolidation: Strategic Effects

In Experiments 1 and 2 consolidation proceeded at the expense of the second item, whereas in Experiments 3 and 4, consolidation was interrupted by the distractor task. These findings imply a strategical modulation of consolidation: ongoing consolidation of the first item is not interrupted by the second item, because participants may not see value in withdrawing attention from it before they obtained a good enough payoff for their effort, or in other words, a stable memory representation. When participants are forced to respond to another task, they no longer can afford to set the consolidation criterion to their leisure: they have to handle the competing task demands.

This finding is against the view that consolidation occurs according to a strict *attentional bottleneck model* (Jolicœur & Dell'Acqua, 1998; Ricker & Hardman, 2017; Stevanovski & Jolicœur, 2007; Zylberberg et al., 2011). To recapitulate, the bottleneck model assumes that consolidation occurs in a ballistic manner: once attention is directed to the consolidation of the incoming sensory information, it cannot be interrupted and has to proceed until it is finished. Only after consolidation is finished attention can be directed to new incoming information. Even though we initially found evidence supporting the attentional bottleneck across Experiments 1-3 (i.e., first-item effect), we showed that consolidation could be interrupted with an attentionally more challenging task. This goes against the assumption of a central bottleneck model of consolidation in working memory.

The observation that memory consolidation is disrupted by a secondary task can be explained by resource-sharing models of attention, one of which are the *central capacity-sharing models* (Lehle & Hübner, 2009; Nieuwenstein & Wyble, 2014; M. Tombu & Jolicœur, 2003). These models assume that attention-demanding processes can occur in parallel as long as they share this limited capacity resource. In our case, this would imply that consolidation and the distractor task would proceed in parallel, but sharing attention between them produces costs for both tasks. It is difficult to predict to which degree performance in each task will drop: resource allocation between the tasks may vary by task demands (Fischer & Plessow, 2015), with for example, a more demanding memory task requiring more attention at the expense of the distractor task, or the other way around. This assumption suggests that consolidation is, at least to some extent, under strategic control. Our results are in line with this account: both the memory task and the processing of the distractor task in Experiments 3 and 4 showed effects of consolidation time. This can be explained by this account by assuming that consolidation of the first item was not interrupted by the distractor task but proceeded at a lower capacity resulting in a lower quality representation, while at the same time the distractor task was also completed at a lower accuracy and with delay.

Another resource sharing model of attention is the *task-switch model* of consolidation (see Nieuwenstein & Wyble, 2014; Ricker et al., 2018). The task-switch account, in contrast to the central-capacity sharing model, assumes that participants can abort consolidation in order to direct their attention towards a new incoming information or task, like the parity-judgment task used in our Experiment 3 and 4. The task-switch account further assumes that participants may not consistently switch their attention to new incoming information. This would predict that participants sometimes abort consolidation to engage into the new task/information and, sometimes, they will finish consolidation of the memory item at the expense of the new incoming task/information. This account could explain again the costs we observed for both tasks.

Our data cannot adjudicate between a resource-sharing model and the task-switch model of consolidation. In order to tease apart predictions from these models, one may need to test whether interference between the distractor task and the memory task depends on parameters that could influence the degree of capacity sharing between them (in line with the capacity sharing model), such as the difficulty of the distractor task or task priorities.

Although it seems like consolidation might be under strategic control, our data suggests that it is still challenging to understand under which circumstances consolidation will be stopped or not. Here, we showed that an attentionally demanding task disrupted ongoing consolidation, whereas manipulations of the memory task setup (recall test order, changes in location of the memory items) did not. Why is ongoing consolidation not interrupted by presentation of the subsequent memory item? One possible explanation is that participants set a criterion for a good enough representation: representations below this criterion may be judged as not worth the effort. This would predict that participants prefer to obtain one fully consolidated representation than two poorly consolidated ones. This hypothesis further predicts that changing processing priorities within the working memory task affects consolidation. Further studies are needed to address to which degree other manipulations – e.g., manipulations that could change processing priorities among memory items – impacts consolidation in visual working memory in line with strategic resource allocation hypothesis proposed here. This will advance our knowledge regarding the degree in which consolidation is under voluntary control.

8.7.2. Verbal Labeling vs. Consolidation Time Benefit?

We argued in the beginning that the consolidation time benefit in Ricker and Hardman (2017) could be due to verbal labeling as these authors did not prevent verbalizations during the consolidation intervals in their experiments. Therefore, our participants in Experiment 1 were instructed to repeatedly say "ba ba ba" aloud during item presentation; yet still, we found a consolidation benefit – ruling out that the consolidation time effect could be fully explained by a verbal labeling effect.

In Experiments 2 to 4, we removed the requirement to perform articulatory suppression. If we contrast the serial input position curves of Experiment 1 (Figure 2A) with the ones obtained in the remaining experiments that did not involve suppression (see Figures 4, 5A, 6A), we can see that their pattern differs. In Experiment 1, where participants performed verbal suppression, the serial position curves were steep with recall error increasing over input/output positions in both consolidation time conditions. In contrast, in the remaining experiments the serial position curves show a much flatter pattern in the long consolidation condition. The latter pattern is similar to the results of Ricker and Hardman (2017), which also obtained a flat line over serial position in their long consolidation condition. This contrast suggests that participants were also benefitting from verbal labeling in addition to consolidation. This is in line with previous results indicating that when participants have sufficient time to label the stimuli, their visual working memory performance improves (Overkott & Souza, 2020; Souza et al., 2020; Souza & Skóra, 2017).

8.7.3. Implications for Working Memory Models

Working memory has a limited capacity for holding memory representations (Cowan, 2010; Oberauer et al., 2016). Many working memory models assume a tight coupling between working memory and attention (for a review see Oberauer, 2019). Short-term consolidation is one of the ways in which attention is assumed to be involved in working memory (Ricker et al., 2018; Souza & Vergauwe, 2018). The present study is in line with the assumption that attention is a limited resource that can be strategically directed towards incoming information for its consolidation in working memory (Bayliss et al., 2015; Kandemir et al., 2017; Nieuwenstein & Wyble, 2014; Oberauer, 2019). Together with the results of Ricker and Hardman (2017), our experiments show dependency between sequentially encoded memory items: if consolidation of a given memoranda is not finished, then a subsequently presented memory item may suffer. Our results also point towards one way to explain dual-task costs in working memory paradigms: processing of distractor information interrupts consolidation of the preceding memoranda in working memory, leading to a less stable representation. These results show important ways by which working memory performance will be constrained by consolidation.

Capacity limitations induced by incomplete consolidation of memory representations have not been readily incorporated in models of visual working memory. The most prominent views assume that either (a) the number of representations that is held in working memory is limited (e.g., Cowan et al., 2013; Fukuda et al., 2010; Zhang & Luck, 2008); (b) a resource defining the quality of the representation is limited (e.g., Bays & Husain, 2008; Brady & Alvarez, 2015; van den Berg et al., 2012); or (c) memory representations are subjected to interference and limited in terms of the bindings formed (Oberauer & Lin, 2017). All these models assume that core capacity limitations are driven by memory maintenance processes, and they do not consider the role of short-term consolidation in the formation of a stable working memory representation. This is probably the case because earlier work suggested that consolidation was finished within a few milliseconds, ca. 50 ms per item (Vogel et al., 2006). This estimate is far below the beneficial effect of time observed here and in other studies (for a review see Ricker et al., 2018).

The effect of consolidation on the stability of a memory representation may be an important factor explaining the susceptibility of this trace to interference and time-based forgetting (Ricker & Cowan, 2014), and hence on explaining capacity limitations in working memory. To date, one of the only models that specify the process of working memory consolidation in a manner that is consistent with the present findings is the eSTST model. This model assumes that consolidation (a) proceeds even after masking, (b) it can occur in parallel, and (c) it can be extended to predict the interruption of consolidation by a secondary task. To recapitulate, the eSTST model assumes that incoming information is activated as a type, which is bound to a token once an activation threshold is reached. This binding process reflects consolidation. Attention serves two functions: it boosts the activation of the type, and suppress the processing of newly incoming information.

This model can explain the findings of our experiments. When the only event following item 1 was the second memory item, consolidation of the first item was not interrupted because item 2 was not attentionally exciting enough to break through the attentional suppression produced by consolidation of the first item. In contrast, in Experiments 3 and 4, the distractor task required an immediate response and this was potentially more exciting, thereby breaking through the attentional suppression generated by consolidation of item 1. Once the distractor task gets enough activation to be processed, it either interrupts ongoing consolidation of item 1 or it shares processing capacity with it, reducing its consolidation.

Working memory models should aim to incorporate the two attentional mechanisms that were suggested by the eSTST model: excitatory attentional boost of relevant incoming information facilitating its consolidation and the suppression of subsequently presented information. These models should consider that during sequential item presentation, items will compete with each other for consolidation. The first item receives priority because its consolidation tends to run through completion, thereby suppressing the reallocation of attention to the consolidation of the following item. The incorporation of these two attentional mechanisms into working memory models will allow for a better prediction of core capacity limitations in working memory. As of now, these models mainly assumed that the limitations are driven by maintenance processes and not by the formation of a stable memory representation. Inclusion of inter-item competition for attention and consolidation, and how distractor processing interrupts these processes will likely provide new ways to understand how working memory capacity is limited.

8.8. Conclusion

Consolidation in working memory is an attentionally demanding process, requiring time to create a stable working memory representation. Contrary to previous claims that this process is all-or-none and cannot be interrupted, we showed that the evidence supporting this assumption rather reflects a functional constraint: presentation of the subsequent item is not attentionally engaging enough to interrupt ongoing consolidation, perhaps because participant are not sufficiently motivated to withdraw attention from the current encoded item before it is fully consolidated. Here, we demonstrated that the insertion of a secondary task demanding an immediate response forced interruption of consolidation mid-way in order to process the distractor task. This revealed that the true nature of consolidation is of a graded process, but one that might be subjected to strategic criterion choices: consolidation can be interrupted if the following stimulus demands immediate attention; otherwise the payoff may not be high enough to make it worth the effort.

9. Verbal Descriptions Improve Visual Working Memory, but Have Limited Impact on Visual Long-Term Memory

Clara Overkott & Alessandra S. Souza

University of Zurich

Status: *under review* in Journal of Experimental Psychology: General (15.06.2020)

Author's contribution

Conceptualization, methodology, and software C.O. and A.S.S.; investigation, formal analysis, visualization, data curation, and writing – original draft preparation, C.O.; writing – review and editing, supervision, project administration, resources, and funding acquisition, A.S.S.

Acknowledgements

This research was support by a grant from the Swiss National Science Foundation (100019_169302) awarded to A. S. Souza. We thank Gian-Luca Gubler and Vanessa Vallesi for helping with data collection. All experiments reported in this paper were pre-registered in the Open Science Framework. Pre-registrations, materials, data, and analysis scripts can be found on: https://osf.io/pn3fh/.

9.1. Abstract

How do verbal descriptions affect visual memory over the short- and long-term? Here we show for the first time that verbal labeling can boost visual memories, but the source of this benefit depends on whether representations are maintained over the short-term in visual working memory, or over the long-term in visual long-term memory. Across three experiments, we contrasted color memory of randomly colored objects when participants labeled (a) the color, (b) the object, or (c) the color-object binding, to retention under an articulatory suppression condition inhibiting labeling. Memory was tested at two time points: after three objects (visual working memory) and at the end of the experiment (visual long-term memory). In Experiment 1, color labeling improved, whereas object labeling impaired, visual working memory in comparison to suppression. Visual long-term memory remained unchanged across conditions. Experiment 2 tested whether this was due to poor overall long-term learning by repeating the colored objects over three successive working memory trials. This increased performance over the short and long-term; yet labeling did not change learning rate over repetitions or delayed memory performance, showing no long-term memory benefit. In Experiment 3, a labeling benefit was observed when the color-object binding was labeled both over the short- and long-term. Mixture modeling indicated that color-labeling benefits in visual working memory resulted from an increase of detailed visual memory, whereas long-term memory benefits accrued from categorical representations. Our findings point to dissociations on the role of language in visual working memory and visual long-term memory.

9.2. Introduction

How do verbal descriptions affect visual memory over the short- and long-term? We may describe the visual information that we need for ongoing processing (e.g., the positions of the cars approaching us while changing lanes), or information that we may need to retain over longer periods (the route we took to arrive at a certain place). Retention of visual information over short and long time-scales are supported by different memory systems. Visual working memory (VWM) keeps visual information available for ongoing cognition. VWM has a limited capacity, and therefore people can only maintain a small amount of information in this system at a given time (Luck & Vogel, 2013; Oberauer et al., 2016). In contrast, visual long-term memory (VLTM) stores large amounts of visual information over long periods of time, varying from several minutes to years, with no upper-limit on how much information can be committed to VLTM (Brady et al., 2008; Konkle et al., 2010a, 2010b).

Verbal labeling has been found to improve VWM (Souza & Skóra, 2017) by increasing the fidelity of the representations stored in this system. In contrast, labeling has been reported to be inconsequential for VLTM (Kelly & Heit, 2017): it produces neither a benefit nor a cost to memory performance over the long-term. What are the reasons for these discrepant findings? The present study aimed to provide a first systematic comparison of how labeling affects visual representations retained in VWM for an immediate task goal, and retained in VLTM for delayed recall.

In the following, we first review how memories are retained over the short-term and long-term in relation to the quantity and quality of the information stored. Next, we describe how labeling has been linked to categorical knowledge, and current hypotheses on how labeling changes visual representations. Finally, we discuss whether there are reasons to suspect that labeling operates differently when memories are stored in VWM vs. VLTM, and then delineate our research aims.

9.2.1. Visual Memories over the Short and Long-term

Memories stored in VWM and VLTM differ in several regards. Research over the past 10 years has demonstrated that visual memories can be described in terms of parameters reflecting its quantity and quality by using mixture models (Zhang & Luck, 2008). In these models, quantity refers to the number of objects accessible for recall, whereas quality refers to the fidelity or precision with which these objects are stored. This approach is commonly applied in the so-called fidelity tasks where participants are required to reproduce, using a continuous scale, one of the features of the memoranda (Prinzmetal et al., 1998b; Wilken & Ma, 2004b; Zhang & Luck, 2008). For example, the participant might be instructed to remember the precise color of a set of real-world objects. At test, the object is presented in grey, and the task is to reproduce the color associated with that object using a continuous color wheel. This task has been used to examine changes in the accessibility and precision of features of a small set of objects maintained in VWM in comparison to the features of hundreds of objects stored in VLTM (Biderman et al., 2019; Brady, Konkle, Gill, et al., 2013). Biderman et al. (2019) showed that both memory precision and the probability of memory retrieval were higher when information was maintained in VWM than in VLTM. This shows that VWM maintenance confers higher accessibility and fidelity to visual representations.

More recently, these mixture models have been extended to incorporate parameters reflecting the contribution of categorical knowledge to memory (Bae et al., 2015; Donkin et al., 2015; Hardman et al., 2017; Persaud & Hemmer, 2016). This is because systematic categorical bias has been uncovered when features are reproduced from perception (Bae et al., 2015), VWM (Donkin et al., 2015; Hardman et al., 2017), and VLTM (Persaud & Hemmer, 2016). In a nutshell, a substantial proportion of responses in fidelity tasks are influenced by the category the memorized feature belongs to (e.g., “red”), rather than the specific feature-value studied (e.g., the specific reddish hue).

Here we will use a categorical-continuous mixture model (Hardman et al., 2017) to probe how conditions prompting and preventing verbal labeling change parameters associated with the storage of categorical and continuous information in VWM and VLTM. Implicitly, categorical knowledge has been related to verbal labeling, whereas continuous information was associated with purely visual memory limitations. In the next section, we present the available evidence for the labeling effects on visual memory over the short- and long-term and how labeling affects categorical and continuous memory parameters.

9.2.2. Labeling vs. Categorical Representations

Although categorical representations are usually assumed to reflect the impact of verbal labeling on visual memory, this assumption has been under-investigated empirically. Recently, Souza and Skóra (2017) manipulated labeling opportunities in a VWM fidelity task: participants studied four sequentially presented colored dots while either (a) labeling the presented colors aloud, or (b) saying “bababa” aloud (a verbal suppression procedure that inhibits labeling). During test, the colors of all four dots were reproduced on a color wheel. The authors observed that color labeling improved recall performance in comparison to suppression. Mixture modeling revealed that color labeling increased the tendency to respond categorically as opposed to guessing. This is in line with the assumption that verbal labels provide categorical information. Surprisingly, labeling also impacted continuous memory by either increasing the proportion of continuous memory responses as opposed to guessing, or the precision of this continuous memory. This effect was interpreted as indicating that the activation of categorical information in VLTM through labeling augmented or protected the continuous representations held in VWM.

In contrast to this labeling benefit in VWM, Kelly and Heit (2017) found that labeling was not unique in improving recognition performance in an episodic VLTM test. In their experiments, participants were presented with a series of colored objects (red or green) and were asked (a) to categorize the colors of objects as being either red or green, or to judge whether (b) they liked the presented object (preference judgement), or (c) the object was living/nonliving (animacy judgement). The specific hue of red or green was irrelevant for the categorization decisions. Episodic VLTM for the specific object color-hue was then assessed in a surprise test at the end. Categorizing the object in regard to its color resulted in a shift towards fewer categorical color responses in the memory test than when participants made preference or animacy judgements. Critically, this did not increase the probability of choosing the correct color. This decrease in categorical responses was also found when foreknowledge of the upcoming VLTM test was given in all conditions. Kelly and Heit (2017) concluded that color labeling reduced categorical bias, but this facilitation was not unique to labeling.

To summarize, these two studies suggest contrasting effects of color labeling on the retention of color in VWM and VLTM. Souza and Skóra (2017) found that labeling benefited VWM by increasing access to both continuous and categorical information. Kelly and Heit (2017) found that labeling reduced categorical bias in a VLTM test, but this did not increase memory for the correct color. These divergent findings may suggest that VWM and VLTM are affected differently by verbal labeling. The caveat here is that these two studies manipulated verbal labeling differently. Kelly and Heit (2017) did not instruct participants to overtly label the colors (they categorized them via keypress), whereas Souza and Skóra (2017) explicitly instructed participants to say the colors aloud. It is unclear whether participants would rely on verbal labels to perform the categorization task used in Kelly and Heit (2017) after a few trials. These divergent findings may therefore reflect differences in the procedure assumed to generate labeling behavior. Another critical difference across these two studies refers to the memory test. In the study of Souza and Skóra (2017), participants reproduced the colors using a continuous color wheel. In the study of Kelly and Heit (2017), participants reported the remembered colors by picking it from a 5-choice alternative set. The latter procedure is limited in the assessment of memory precision and might therefore reduce the chance of measuring a labeling benefit. Accordingly, before we can conclude that labeling affects VWM and VLTM differently, these two systems need to be compared under equivalent conditions. This will be one of the main goals of the present study.

Before we move to the empirical work, it is important to understand the proposed mechanisms by which labeling can influence visual memories. Several hypotheses have been raised, and we will review them in the following section.

9.2.3. Hypotheses of the Labeling Effect

Here, we will discuss five hypotheses that make differential predictions regarding how labeling affects storage of categorical and continuous representations. It is worth noting that none of these hypotheses make differential predictions regarding VWM vs. VLTM, and most of them have received support from research evaluating either of these memory systems. This is probably the case because the effects of labeling on VWM and LTM have not been put in direct comparison before.

9.2.3.1. 1. Verbal Recoding

The *verbal recoding hypothesis* (Souza & Skóra, 2017) assumes that during encoding verbal labeling creates a verbal trace at the expense of the visual information. This hypothesis has also been referred to as "*label distorting memory*" (Kelly & Heit, 2017). For example, labeling the picture of a light-blue shoe as "blue" creates a verbal trace of "blue" whereas the visual details about the specific hue (e.g., shade of light blue) are lost. This hypothesis therefore predicts a cost of labeling for detailed visual memory.

Evidence for the verbal recoding hypothesis stems from the *verbal overshadowing* effect in VLTM (Lupyan, 2008; Schooler & Engstler-Schooler, 1990). For example, Lupyan (2008) asked participants to label objects as belonging to either one of two categories (e.g. chair vs. lamp) and observed worse long-term recognition performance compared to a preference rating condition. Moreover, Schooler and Engstler-Schooler (1990) showed that a color category label interfered with recalling the correct color hue in VLTM, as this specific information was no longer available in memory (see also, Alogna et al., 2014; Brandimonte et al., 1997).

9.2.3.2. 2. Dual Trace

The *dual-trace hypothesis* (Souza & Skóra, 2017) assumes that labeling builds two memory traces: a verbal trace based on the verbal label that was assigned to the object and a visual trace of the object itself. This stands in contrast to the verbal recording hypothesis, where labeling is assumed to generate only one verbal (categorical) trace. This hypothesis predicts that labels help memory by providing an additional source of categorical information, without changing the retention of the visual trace. This assumption is exemplified in the modeling implemented by Donkin et al. (2015): they included verbal labeling as a further component into a mixture model estimating the quantity and quality of VWM representations. Their modeling showed that the inclusion of this parameter better predicted their VWM data, because some responses seemed to have been guided by information provided by the label. Their modeling, however, does not assume that labeling induces any change in the visual trace.

Further evidence for the dual-trace hypothesis was found in VLTM studies showing that the verbal overshadowing effect could be modulated or even reversed (Brandimonte et al., 1997; Brown et al., 2014). For example, Brown et al. (2014) asked participants to learn easy-to-label and hard-to-label pictures, with the assumption that participants would covertly label the easy-to-label pictures. Then, participants were asked to either provide a detailed description of the learned feature or do a filler task. The final memory test was meant to either favor retrieval of featural or global information of the object. These authors found that covert verbal labeling of the easy-to-label pictures impaired VLTM performance, as would be predicted by the verbal overshadowing effect (see also, Brandimonte et al., 1992). However, a detailed description of the feature benefitted VLTM performance in a featural memory test. This provides evidence that the verbal overshadowing effect for labels can be reversed with feature descriptions that match the final memory test. This finding challenges the verbal recording hypothesis by showing that participants may have both the visual and the verbal traces accessible.

9.2.3.3. 3. Distinctiveness

The third hypothesis proposes that verbal labels make memory representations more distinct (Blanco & Gureckis, 2013; Kelly & Heit, 2017; Richler et al., 2013; Souza & Skóra, 2017). This *distinctiveness hypothesis* assumes that a label serves as an additional retrieval cue to the memory object or as a cue to augment encoding specificity (Blanco & Gureckis, 2013; Richler et al., 2011; Tulving & Thomson, 1973), thereby facilitating memory retrieval. Critically, if labels simply provide a distinctive cue to memory, it should not matter what type of label is used, as long as it provides a unique means to access the visual trace.

In verbal studies, a distinctiveness effect has been obtained in the comparison of memory for words read aloud vs. silently during study (MacLeod, 2010; MacLeod et al., 2010; Ozubko & MacLeod, 2010). For VLTM, Richler et al. provided some evidence for a distinctiveness benefit: they presented exemplars from either unique categories or exemplars sampled from only two categories. They showed that vocally labeling the unique categories during study yielded similar memory performance as a preference rating task. In contrast, the two category labels impaired memory performance. Additionally, preference ratings using a 5-point scale during encoding provided more distinctiveness and presumably deeper processing than the labeling of the memory items with two categories (Blanco & Gureckis, 2013). These studies suggest that the uniqueness of a category label is essential for a distinctiveness benefit: the more unique, the better. Souza and Skóra (2017) also tested whether distinct labels could improve VWM for colors. They instructed participants to label the presentation order of a sequence of four colors (e.g., first, second, third, and forth) under the assumption that these labels would increase distinctiveness in comparison to a condition with articulatory suppression. However, labeling their serial position did not provide any advantage.

9.2.3.4. 4. Activation of Categorical VLTM

The *activation of categorical VLTM hypothesis* (Souza & Skóra, 2017), which is based on the *label-feedback hypothesis* (Lupyan, 2012), assumes that verbal labels activate categorical knowledge in VLTM. In this case, two visual traces are produced: one from visually encoding the object and the other is the VLTM representation of the category activated by the verbal label. Activation of the visual categorical representation may allow data compression (see also, Brady et al., 2009): instead of storing all of the details regarding the visual object, the memory trace may represent deviations in relation to the category, thereby reducing memory load. Accordingly, this hypothesis predicts a labeling benefit with more visual details being stored in memory. Evidence for a labeling benefit of this sort has been obtained by Souza & Skóra (2017): they showed that verbally labeling colors improved VWM compared to a suppression condition due to increases in categorical and continuous memory.

Further support for this hypothesis stems from studies finding that labels more efficiently cued the category (e.g., dog) of an object than non-verbal stimuli (e.g., a barking sound), thereby facilitating categorization and perceptual decisions (Boutonnet & Lupyan, 2015; Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012).

9.2.3.5. 5. Cue to Focus Attention

Labels can also be viewed as a *cue to focus attention* in certain aspects of the visual object (Kelly & Heit, 2017). This hypothesis predicts that labeling may only be useful if it guides attention to relevant features, whereas it may be costly if it guides attention to irrelevant features. Critically, if attention is guided to the labeled feature irrespective of labeling, then labeling should be inconsequential. Kelly and Heit (2017) found that color labeling during study reduced color bias towards the color prototype in a surprise VLTM recognition test in comparison to conditions that required an animacy judgment or preference rating during study. They argued that this occurred because the label guided attention to the relevant feature during study for the later memory test. When participants were informed about the relevant feature for the test before study, the advantage of color labeling vanished.

9.2.4. The Present Study

The main goal of the present study was to examine the impact of verbal labeling on both VWM and VLTM using a color fidelity task. To the best of our knowledge, no previous study considered the impact of labeling concurrently on these two memory systems. Although the hypotheses of the labeling effect do not make differential predictions for retention over short and long timescales, there is empirical reason to suspect that labeling affects VWM and VLTM differently. For example, whereas Souza and Skóra (2017) found a benefit of color labeling to retention of visual details in VWM, Kelly and Heit (2017) found neither benefits nor costs of color labeling in a VLTM test. These findings are difficult to directly compare, however, because their experimental set-up differed in many regards. Accordingly, it is not clear to what degree their contradicting results reflects aspects of the experimental procedure vs. true differences on the creation of visual memory representations to be used for ongoing cognition (e.g., in VWM) vs. for later recall (e.g., in VLTM). Here we designed a task to measure both memory systems using the same type of overt labeling manipulation and task requirements. This allowed us to directly examine how verbal labels influence the creation of memory representations to be accessed over the short- and long-term and to test predictions of the labeling hypotheses delineated above.

Given that the labeling hypotheses do not differentiate between VWM and VLTM storage, this leads to the expectation that whatever mechanism operates over the short-term should also affect performance over the long-term. Our experiments provide a unique opportunity to address whether this is indeed the case. If the effect of labeling differs between VWM and VLTM, this would require a revision of the labeling hypotheses and would support the separation of these two memory systems as independent of each other (Brady et al., 2011).

The general procedure of our experiments was as follows. We implemented two phases: a VWM phase containing the labeling manipulations, followed by a final delayed memory test that comprised our VLTM phase. In the VWM phase, participants completed several trials of a continuous color fidelity task. Trials consisted of the sequential presentation of three colored objects. To assess the effect of verbal labeling on memory, participants were instructed to either (a) label the color (Experiments 1 and 2), (b) label the object (Experiment 1), or (c) label the color-object combination (Experiment 3). As a control condition in all experiments, participants also performed the task while saying “bababa” aloud (suppression) thereby inhibiting the use of verbal labeling.

At the VWM test, participants were tested on their memory for the colors of all three objects: they were shown the object in grey as a retrieval cue and they were asked to reproduce its color using a color wheel. After the end of the VWM phase, participants were asked to reproduce the color of all objects studied again (VLTM phase). Our goal was to examine whether retrieval of an object’s color in the VLTM test would vary depending on the labeling manipulations implemented during the VWM phase. This allowed us to test whether labeling would affect memory representations similarly when they were retrieved from VWM and from VLTM.

To foreshadow our results, we found a benefit of labeling the color and a cost of labeling the objects for the retention of color-object combinations in VWM in Experiment 1. There was no effect of labeling on VLTM, independently of whether participants had foreknowledge about the VLTM test (Experiment 1b) or not (Experiment 1a). However, overall performance in the VLTM test was quite poor. To improve VLTM learning, in Experiment 2, each trial of the VWM phase was repeated three times to increase long-term learning. Additionally, participants were only required to label the colors or to perform suppression (the object labeling condition was dropped). Across the three repetitions, performance improved in the VWM test thereby showing a learning effect. There was a color labeling benefit in VWM for the very first presentation of the color-object binding, but this benefit vanished over the course of the repetitions. Although performance improved overall in the final test, Experiment 2 showed no labeling effect in VLTM replicating Experiment 1. In Experiment 3, participants were asked to label the color-object combinations (instead of only the color or only the object) and this was contrasted to suppression. For the first time across our series of experiments, we showed a labeling benefit in both VWM and in VLTM.

Overall, we found evidence for a dissociation of the labeling benefit between the short-term and the long-term. Modeling further showed that labeling benefited continuous memory over the short-term, whereas this benefit was categorical in the long-term. This indicates that (a) the labeling benefit has different sources in VWM and VLTM or (b) that labels are either not retained over the long-term or they do not survive the interference that accumulates in VLTM.

9.3. Experiment 1

The goal of Experiment 1 was to investigate whether the beneficial effect of color labeling in VWM would translate into better color memory in VLTM. In addition, we included an object labeling condition that allowed us to further distinguish between the predictions of the labeling effect.

In the present experiment, participants were asked to: (1) say “bababa” aloud thereby inhibiting labeling, (2) label the color, or (3) the shape of visual objects during the VWM phase. At the end of the study, they were then tested again on the same visual objects in a delayed memory test (VLTM phase). The memory test in the VWM and VLTM phases required participants to reproduce colors using a continuous color wheel. The use of a continuous color test allowed us to assess how labeling affected the storage of continuous and categorical information in both memory systems using a mixture modeling approach.

The five hypotheses of the labeling effect make differential predictions for the data of Experiment 1, which are summarized in Table 1. (1) The *label recording* hypothesis predicts a labeling cost compared to the suppression baseline. This cost should be reflected on memory precision in the color labeling condition as the label replaces the fine-grained detail of the color hue. In the object labeling condition, in contrast, it should be reflected on the accessibility of the memory representation because the object’s name would overshadow the color information. (2) The *dual-trace* hypothesis predicts an increase in categorical responding as a function of color labeling with no change in continuous information. Object labeling should have no effect on memory performance, because this label lacks in providing information to improve color recall; (3) The *distinctiveness* hypothesis predicts that labeling should increase the chance of recalling the visual information, and this increase should be larger for object than color labeling given that object labels provide a more unique cue to the memory representation. (4) The *activation of categorical VLTM* hypothesis predicts that labeling yields a benefit not only to categorical but also continuous visual information. This benefit should only be observed to color labels, because they are the only ones that activate the relevant categories to the memory test. Lastly, (5) the *cue to focus attention* predicts that color labeling should be inconsequential since participants were already fully aware that color information was the relevant feature; object labeling, in contrast, should lead to a cost because it draws attention away from the relevant feature for the test.

Table 1

Summary of Predictions of the Labeling Hypotheses to the Data of Experiment 1

Hypothesis	Color Labeling	Object Labeling
1. Label Recoding	↓ Memory Precision	↓ Memory Accessibility
2. Dual trace	↑ Categorical Responses	=
3. Distinctiveness	↑ Accessibility	↑↑ Accessibility
4. Activation of categorical VLTM	↑ Continuous Memory	=
5. Cue to Focus Attention	=	↓ Memory

We ran two experimental versions. In Experiment 1a, participants were not informed about the VLTM phase, whereas in Experiment 1b, participants were informed about the VLTM phase at the beginning of the study. Our reasoning to disclose the occurrence of the VLTM test in Experiment 1b was to motivate participants to try to remember the objects over the long-term, thereby possibly increasing VLTM performance.

The research questions, method, and statistical hypotheses for Experiment 1a were preregistered and can be found at: https://osf.io/wru4z/. Note that our preregistration was only concerned with differences between VWM and VLTM with regards to the effect of labeling. Predictions regarding the hypothesis of the labeling effect were not preregistered. Experiment 1b was a replication with just one minor modification in the instruction and was not preregistered. We maintained the same pre-registered analysis plan for both experiments.

9.3.1. Methods

9.3.1.1. Participants

Fifty-seven students of the University of Zurich participated in this experiment. Only participants with German (or Swiss-German) mother tongue, aged between 18-35 years, and reporting normal color vision or corrected-to-normal visual acuity could take part in the experiment. Participants signed an informed consent prior to the study and were debriefed at the end. The experimental protocol was in accordance with the guidelines of the Institutional Review Board, and it did not require special approval.

The first 30 participants took part in Experiment 1a (*M* = 27.73, *SD* = 3.74, 23 women) and the next 27 participants were assigned to Experiment 1b (*M* = 23.19, *SD* = 3.56, 16 women). Six participants were excluded from Experiment 1a as they failed to follow the labeling instructions[4], resulting in a final data set of 24 participants. Three participants were excluded from Experiment 1b[5], resulting in a total of 24 participants. As detailed in our preregistration, we aimed to collect data of at least 30 participants in Experiment 1a, and we were going to adjust the sample size based on the evidence obtained for or against our hypotheses. The final sample size in these experiments was sufficient to provide substantial evidence to answer our research questions, hence we stopped data-collection as reported in the preregistration.

9.3.1.2. Materials

All experiments were programmed in MATLAB (2010b for Experiment 1; 2016b for Experiments 2 and 3) using the Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997). Nameable clip-art pictures served as stimuli objects, which were taken from Sutterer and Awh (2016). The objects were colored in one out of 360 colors that varied along a continuous color wheel (Zhang & Luck, 2008), defined in the CIELAB color

---

[4] Four of these participants did not follow the instruction to switch between labeling conditions on several occasions and remained labeling the wrong condition for the entire block (e.g. they continued labeling the color instead of the object), and two did not label at all.

[5] One participant verbalized only on some trials, one participant confused the labeling conditions, and one participant labeled the fixation cross instead of the objects.

space with L= 70, a= 20, b= 38, and a radius of 60. The colored objects were presented against a grey background (RGB 128 128 128). Participants saw each object once. The color-object combinations (hereafter referred as bindings) were randomly selected for every participant.

9.3.1.3. Procedure

VWM phase. Each VWM trial started with a 1000 ms fixation cross in white (RGB 255 255 255) in the center of the screen. Thereafter, a sequence of three objects was presented. Each object remained onscreen for 250 ms, followed by a 1000 ms blank inter-object interval, providing time for labeling (see Figure 1A). To investigate how labeling influences VWM and VLTM we introduced three labeling conditions during the study phase: (a) label the color (e.g., “red”), (b) label the object (e.g., “heart”), or (c) suppression (e.g., “bababa”). These labeling instructions appeared at the beginning of each trial to remind participants of the current condition. Participants were asked to self-initiate each trial by pressing the space bar. They were further instructed to wear a headset and their verbal responses were recorded for offline check. The labeling conditions were completed in short blocks of 8 trials, and blocks of different conditions alternated (e.g., suppression-color-object-suppression-color-object). The order of the conditions was counterbalanced across participants. Each condition contained 3 practice trials and 32 experimental trials. The practice trials were completed right before the first block of this condition. Overall, there were 105 objects per condition (including practice trials), and 315 objects in total.

In the VWM test, all three objects were tested in random order (see Figure 1B). The memory test phase was initiated by the presentation of a dark-grey wheel (RGB 96 96 96) around the tested object, which was presented in light grey (RGB 160 160 160). Once participants started moving the mouse along the grey wheel, the color of the probe changed. Participants were asked to adjust the color of the probe to the one they remembered for this object. Once participants right-clicked on the mouse, their color selection was registered, and the next object was presented.

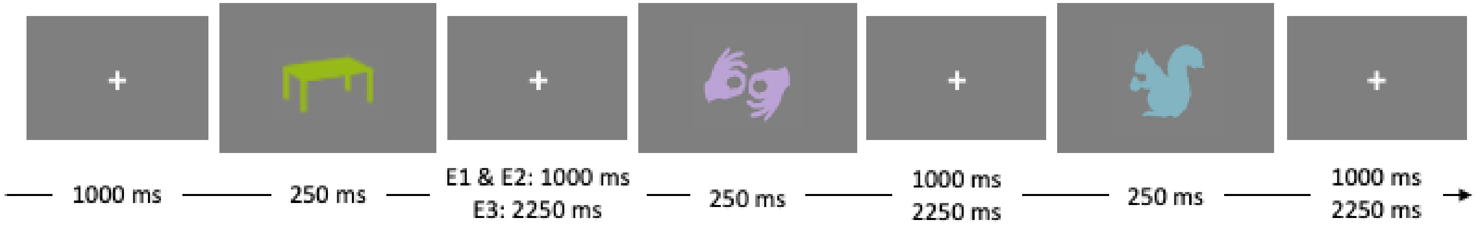
VLTM phase. At the very end of the VWM phase, participants were instructed to leave the experimental room and take a short break for about 5 minutes. During the break, they were offered some sweets (e.g., chocolate). After the break, participants underwent the VLTM test phase. This test phase matched the procedure of the VWM test. In Experiment 1a, participants were not aware of the VLTM test, and hence the delayed test came as a surprise. In contrast, participants in Experiment 1b were informed prior to the start of the experiment that they would have to recall all of the presented objects at a second stage of the experiment, and they were encouraged to try to retain the objects for a longer duration in memory. In both experiments, participants were tested for all the objects from the VWM phase, excluding the practice trials. In total, 288 objects were tested in the VLTM phase, 96 from each labeling condition.

9.3.1.4. Data Analysis

Verbal Labeling Output. We recorded the verbal responses during the study phase. Color labeling responses were coded to assess the variety of labels applied to the colors, and to estimate the color range to which these labels referred to in each experiment. This information was then used to inform our mixture modeling about participants' color categories in each experiment, following the procedure used by Souza and Skóra (2017).

Figure 1

Illustration of the Flow of Events in the Trials of All Experiments Reported Here.

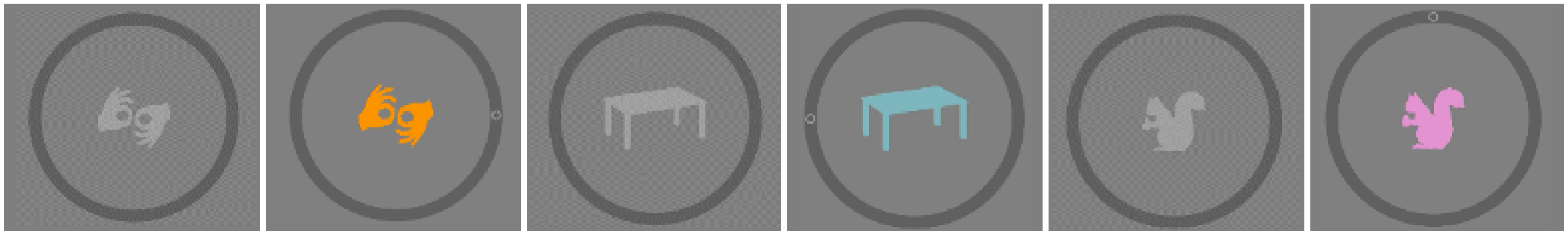


A

Experiments 1 & 2	Color Labeling	*«green»*	*«violet»*	*«blue»*
Experiment 1a & 1b	Object Labeling	*«table»*	*«hands»*	*«squirrel»*
Experiment 3	Color + Object Labeling	*«green table»*	*«violet hands»*	*«blue squirrel»*
All Experiments	Suppression	*«ba ba ba»*	*«ba ba ba»*	*«ba ba ba»*

Random Recall Test

B

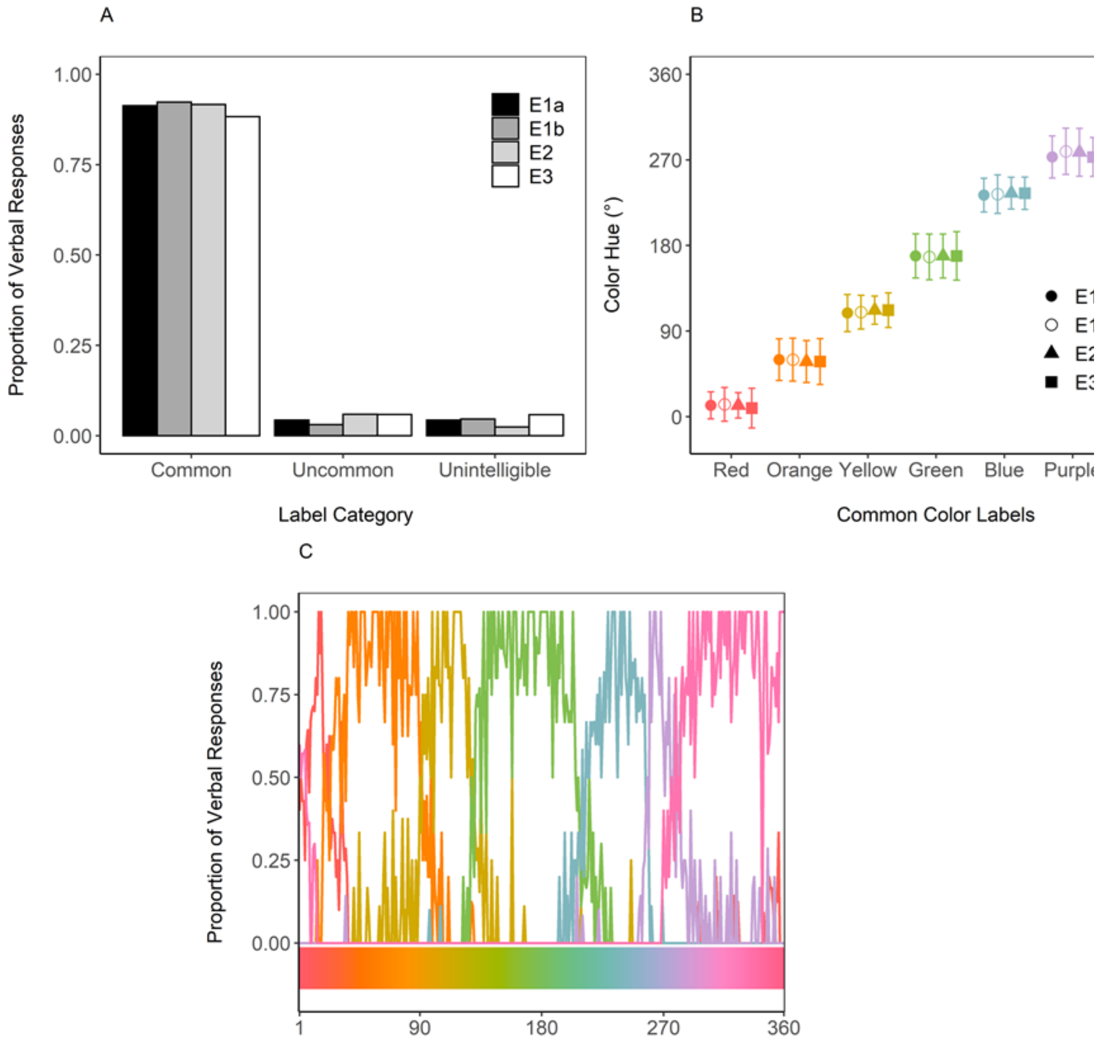


Note. Panel A exemplifies the flow of one trial with examples of the actual objects used for all experiments. Below each object, the applied labeling conditions are illustrated. Panel B shows the random recall test procedure of this VWM trial. Participants first saw a probe in grey. Once participants moved the mouse along the wheel the object's color changed. For VLTM, all objects were tested in the same manner.

Participants used a total of 20 different color labels in Experiments 1a and 1b, 50 in Experiment 2, and 76 in Experiment 3. Similarly to Souza and Skóra (2017), the majority of the color labels belonged to a set of basic color categories (e.g., red, orange, yellow, green, blue, purple, and pink) across all our reported experiments. Figure 2A shows the proportion of verbal responses that fell within these seven color categories (hereafter referred here as common category), as opposed to the usage of more uncommon labels (e.g., turquoise, yellow-green, dark orange, blueish), or unintelligible responses. This figure shows that although various labels were used overall, these uncommon responses were of very low frequency. Figure 2C presents the proportion of times of the seven basic color labels were used (across all participants) to refer to the 360 colors in the color wheel. This led to seven bell-shaped distributions across the continuous color space. The bell-shape of these distributions resembles a normal distribution, and hence we fitted a normal distribution for circular space (e.g., a von Mises distribution) to this data. The von Mises distribution is described by the mean and the standard deviation. These parameters can be taken to define the center of the color category and the variance around it. Figure 2B shows the center of each color category (dot) and the standard deviation of the color categories as estimated by the von Mises fitted to the verbal responses in each experiment.

Figure 2

Analysis of the Color Labels used by the Participants Across All Experiments



Note. Panel A shows the proportion of color labels grouped by the common, uncommon, and unintelligible label categories for Experiments 1-3. Panel B shows the average color for which a given label was assigned and the standard deviation of colors to which the label was applied. These parameters were estimated by a von Mises fitted to the distribution of color label responses over the color space in all experiments. Panel C shows the proportion of times one of the seven common color labels was used to refer to a given color on the wheel (as shown in the x-axis) in Experiment 1a. A proportion of 1 indicates that the x color on the wheel was labeled with the same label by all participants. The lower the proportion, the less often participants used that label to refer to that given color. Each color term is represented by the line with its prototypical color.

Recall. Recall was assessed by calculating the deviation between the given response and the true color value of the studied object in degrees, ranging from +180 to -180 degrees. The absolute value of the deviation can be taken as a model-free index of performance, which we will refer here to as recall error. Our first set of analyses focused on differences between labeling conditions with regards to recall error in the VWM and VLTM tests. We conducted Bayesian Inference statistics because this approach is known to have several statistical advantages over frequentist statistics that rely on p-value significance testing. For example, p-values have the tendency to overstate evidence in favor of the alternative hypothesis (Wetzels et al., 2011). In contrast to p-values, Bayesian inference quantifies the evidence for one hypothesis over the other. One commonly employed measure is the Bayes Factor (BF). The BF is the strength of evidence for one hypothesis (e.g., the alternative) over another hypothesis (e.g., the Null), given the observed data. The advantage of a Bayesian approach is that one can gauge evidence for the alternative and for the null hypothesis. A BF_{10} (e.g., the likelihood of the alternative hypothesis, H1, over the null hypothesis, H0) above 1 yields evidence in support of H1, whereas a BF_{10} below 1 provides evidence in support of H0. BFs should be interpreted as a continuous index of the strength of evidence in the data in support of one model over the other, and provides the factor by which the ratio of our prior beliefs should be updated in light of the data. For example, a $BF_{10} = 10$ indicates that the alternative hypothesis is 10 times more likely than the null hypothesis, given the data. Usually, BFs > 3 are considered as providing substantial evidence for one hypothesis over the other, whereas a $BF \geq 10$ is usually considered as strong evidence. We computed the BFs as stated in Rouder, Morey, Speckman and Province (2012) using the default settings of the BayesFactor package (Morey & Rouder, 2015) implemented in R (R Core Team, 2014).

Experiment E1a and E1b were within-subject designs with 2 (memory test: VWM, VLTM) x 3 (labeling condition: color, object, suppression) factors. These two factors were set as fixed predictors in the BANOVA, and the subject factor was treated as random effect. To compute a BF, the believed probabilities of the parameter distributions, also known as a-priori beliefs or priors need to be set judiciously and computationally convenient (Rouder et al., 2012). The Bayes Factor package provides three default priors that are within a reasonable range. Here, the BFs were computed with the most conservative default prior of $\sqrt{2}/2$. The chosen prior reflects our beliefs about the likelihood of an effect in our experiment. Rouder, Morey, Verhagen, Swagman and Wagenmakers (2017) showed that the prior specification matters, but it does not greatly change the evidence within a reasonable range of prior specifications, such as the range between 0.2 and 1 (which is within the rage of our prior specification). The higher the BF, the less influential the prior is.

In the pre-registrations we stated that we aimed to report BFs ≥ 10 for or against the alternative hypothesis for the main effects and the interactions of interest in the model, which is usually considered as strong evidence.

Categorical-Continuous Mixture Modeling. We modeled the responses in our task using the Bayesian hierarchical categorical-continuous mixture model of Hardman et al. (2017). The model assumes that responses are either informed by memory (P^M) or reflect guessing (1- P^M). Responses informed by memory could reflect continuous (P^O) or categorical (1 - P^O) information about the visual stimulus. Continuous information allows for a fine-grained response that varies linearly with the studied feature. The continuous response can be more or less fine-grained – which reflects the continuous imprecision (σ^O) of the memory representation. In contrast, categorical responses cluster around some canonical values (the category mean) along the feature space. The model further assumes two sources of guessing: guessing could either be categorical, when participants randomly guess prototypical colors, captured by the parameter P^{AG}, or continuous, when guesses are uniformly distributed along the feature space (1 - P^{AG}). In this mixture model, every category has a mean and standard deviation, which can be estimated freely by the model if no prior knowledge about the participants' categories is given. In the following experiments, we fixed the category means using the information extracted from the labeling responses (see Figure

2B), similarly to the approach used by Souza and Skóra (2017) .[6] Further parameters of the model are the category imprecision (how precise is the categorical response) and the categorical selectivity which estimates how selectively colors are assigned to a category.

For all analysis reported in this paper, we fitted the between-item model of the CatContModel package (Hardman, 2016) implemented in R. The between-item model variant assumes that both, categorical and continuous information relative to a stimulus can be hold in memory at the same time. At the point of response selection, however, the response is based on either the categorical or the continuous information, but not both. This model variant has previously been reported to have better model fit (Hardman et al., 2017; Souza & Skóra, 2017) than the alternative variant assuming that responses reflect a combination of both continuous and categorical information. Hierarchical models view the parameters of individual participants in a given condition as samples from a population-level distribution. The parameter values and distributional probabilities were determined through Markov chain Monte Carlo (MCMC) sampling techniques.

For each experiment, we fitted a model that allowed the three main parameters P^M, P^O, and σ^O in the model to vary across experimental conditions. Then, we assessed the posterior estimates of the parameters of the model with regards to the effects of our manipulations. Our main interest was to assess how labeling changed the probability of responses informed by categorical as opposed to continuous information, and the continuous imprecision of the memory representation across both the VWM test and the VLTM test. To assess the reliance on continuous information, P^M needed to be multiplied by P^O. To assess

[6] We also fitted the model allowing free estimates of the color categories across all experiments, which can be found on the OSF, and the results of these model were fairly in line with ones reported here (but see Experiment 3).

reliance on categorical information the equation is as follows: P^M x ($1-P^O$). The continuous imprecision parameter (σ^O) was used as outputted by the model [7].

9.3.2. Results

9.3.2.1. Recall Error

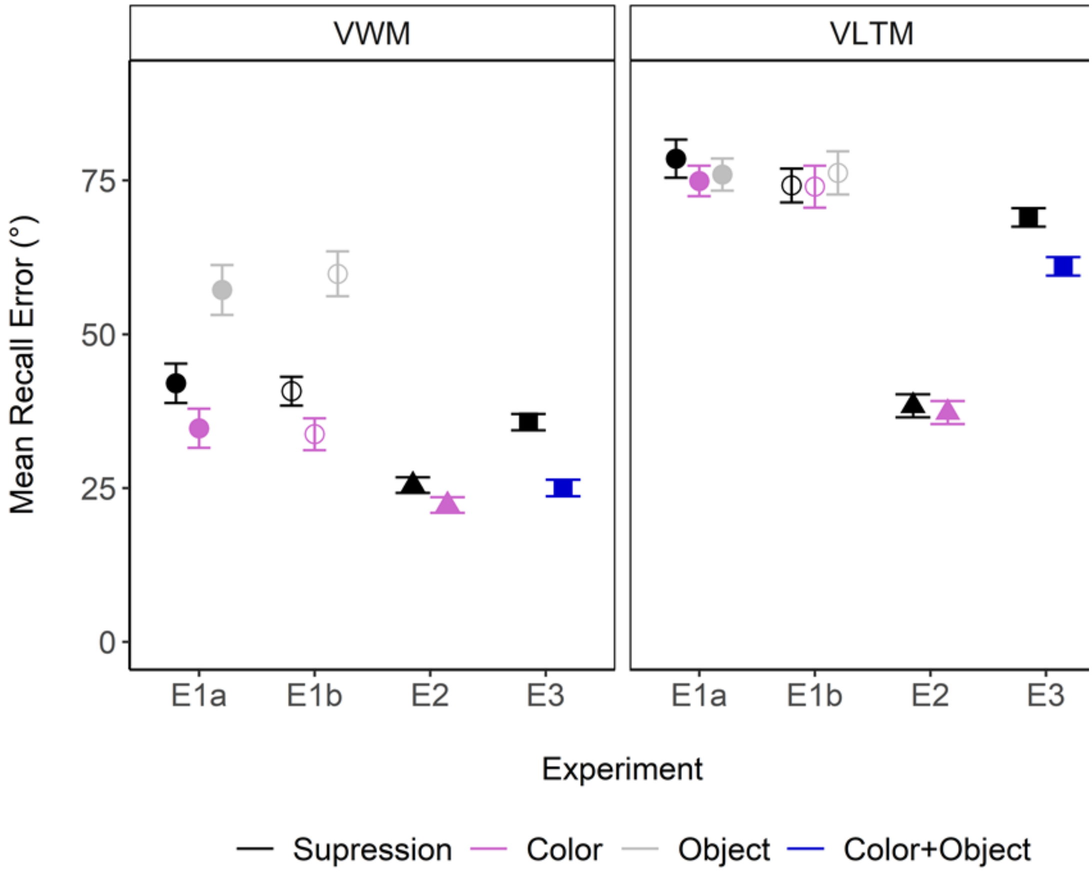
In the preregistration we mentioned to check the residuals of recall error for the assumption of a normal distribution by looking at the QQ plot of the residuals. To check the homogeneity of variance distribution for the recall error analysis, we calculated the variance of the mean recall error for every participant in every condition. The difference in variance in groups was below 4, which is the threshold for assumption violation (Tabachnick & Fidell, 2013). This information can be found on the OSF.

Mean recall error as a function of labeling condition for the two memory tests is presented in Figure 3. Recall error was smaller in the VWM test than in the VLTM test, reflecting better performance in the former. VWM performance improved when participants labeled the colors, whereas it decreased when participants labeled the objects, compared to the suppression condition. Labeling had no discernable effect on VLTM performance.

[7] In the preregistration, we mentioned that we would transform these values into the commonly used capacity K (Cowan, 2001), which requires multiplying the parameters by memory set-size. This produces, however, a very different scale range for VWM (0-3 items) and VLTM (0 to hundreds of items). We decided therefore to keep parameters in the scale from 0-1 for both memory systems. This decision is inconsequential for the assessment of the presence of effects.

Figure 3

Mean Recall Error in Degrees across All Experiments for the VWM and VLTM Tests



Note. The mean error in VWM for Experiment 2 averaged across the three repetitions of the same object. Error bars represent the 95% within-subjects confidence interval.

In line with our preregistered analysis, we conducted a Bayesian ANOVA on the data of Experiments 1a and 1b. Table 2 presents the BFs of all tested models against the Null. The model with the highest BF against the Null is the best model. Our preregistered analysis was mainly concerned with the evidence for an interaction between labeling and memory test. The best model of the data in Experiments 1a and 1b included the effects of labeling condition, memory test, and their interaction. To assess the evidence for the inclusion of the interaction in the best model, we computed the ratio of the best model against the model with only the two main effects. As shown in Table 2, there was overwhelming evidence for the inclusion of the interaction between labeling condition and memory test in the best model of both experiments, indicating that labeling impacted VWM and VLTM differently.

Table 2

*Relative Likelihood of Models with Different Fixed Effects Over the Null Model (BF_{10}) and Relative Likelihood of the Best Model (*e.g., *the One with Higher Likelihood Over the Null) Over the Alternative Model Specified in Each Row (BF_{Best}/BF_{Mrow})*

Exp.	Model n°	Included Fixed Effects: Labeling condition	Memory Test	Labeling x Test	BF_{10}	BF_{Best}/BF_{Mrow}
1a	**1**	✓	✓	✓	$\mathbf{1.17 \times 10^{47}}$	**1**
	2	✓	✓	---	2.68×10^{39}	4.38×10^{7}
	3	✓	---	---	3.06	3.83×10^{46}
	4	---	✓	---	6.94×10^{33}	1.69×10^{13}
1b	**1**	✓	✓	✓	$\mathbf{3.38 \times 10^{46}}$	**1**
	2	✓	✓	---	2.73×10^{37}	1.24×10^{9}
	3	✓	---	---	37.97	8.91×10^{44}
	4	---	✓	---	6.32×10^{28}	5.35×10^{17}
2	1	✓	✓	✓	3.36×10^{26}	3.05
	2	✓	✓	**---**	$\mathbf{1.02 \times 10^{27}}$	**1**
	3	✓	---	---	0.41	2.49×10^{27}
	4	---	✓	---	7.92×10^{26}	1.29
3	1	✓	✓	✓	2.76×10^{86}	1.78
	2	✓	✓	**---**	$\mathbf{4.91 \times 10^{86}}$	**1**
	3	✓	---	---	45.52	1.08×10^{85}
	4	---	✓	---	4.89×10^{69}	1.00×10^{17}

Note. ✓ = effect included in the model. The model with the highest BF against the Null (best model) is printed in bold.

As a follow-up analysis on the interaction [8], we assessed the impact of color and object labeling for the VWM and VLTM test separately. We computed Bayesian *t*-tests to compare both labeling conditions to the suppression condition. For VWM, the difference between color labeling and suppression yielded a $BF_{10} = 36.04$ in Experiment 1a and a $BF_{10} = 1.12 \times 10^{3}$ in Experiment 1b, indicating strong support for a color labeling benefit. In

[8] This set of analyses was not preregistered.

contrast, the difference between object labeling and suppression yielded overwhelming evidence for an object labeling cost in both experiments (Exp. 1a: $BF_{10} = 3.54 \times 10^3$; Exp. 1b: $BF_{10} = 1.31 \times 10^6$). For VLTM, there was ambiguous to substantial evidence for an absence of a color labeling benefit with a $BF_{10} = 0.91$ ($BF_{01} = 1.10$) in Experiment 1a and $BF_{10} = 0.22$ ($BF_{01} = 4.64$) in Experiment 1b. Likewise, there was ambiguous to substantial evidence against an object labeling cost in VLTM: $BF_{10} = 0.43$ ($BF_{01} = 2.30$) in Experiment 1a and a $BF_{10} = 0.33$ ($BF_{01} = 3.04$) in Experiment 1b.

In sum, these results indicate that the color labeling benefit and object labeling cost found in VWM were no longer credible when memory was tested over a delay. Overt verbal labeling clearly affects VWM, but seems to neither benefit nor harm VLTM – with the latter being more evident in Experiment 1b, in which participants were aware of the upcoming VLTM test.

9.3.2.2. Categorical-Continuous Mixture Modeling

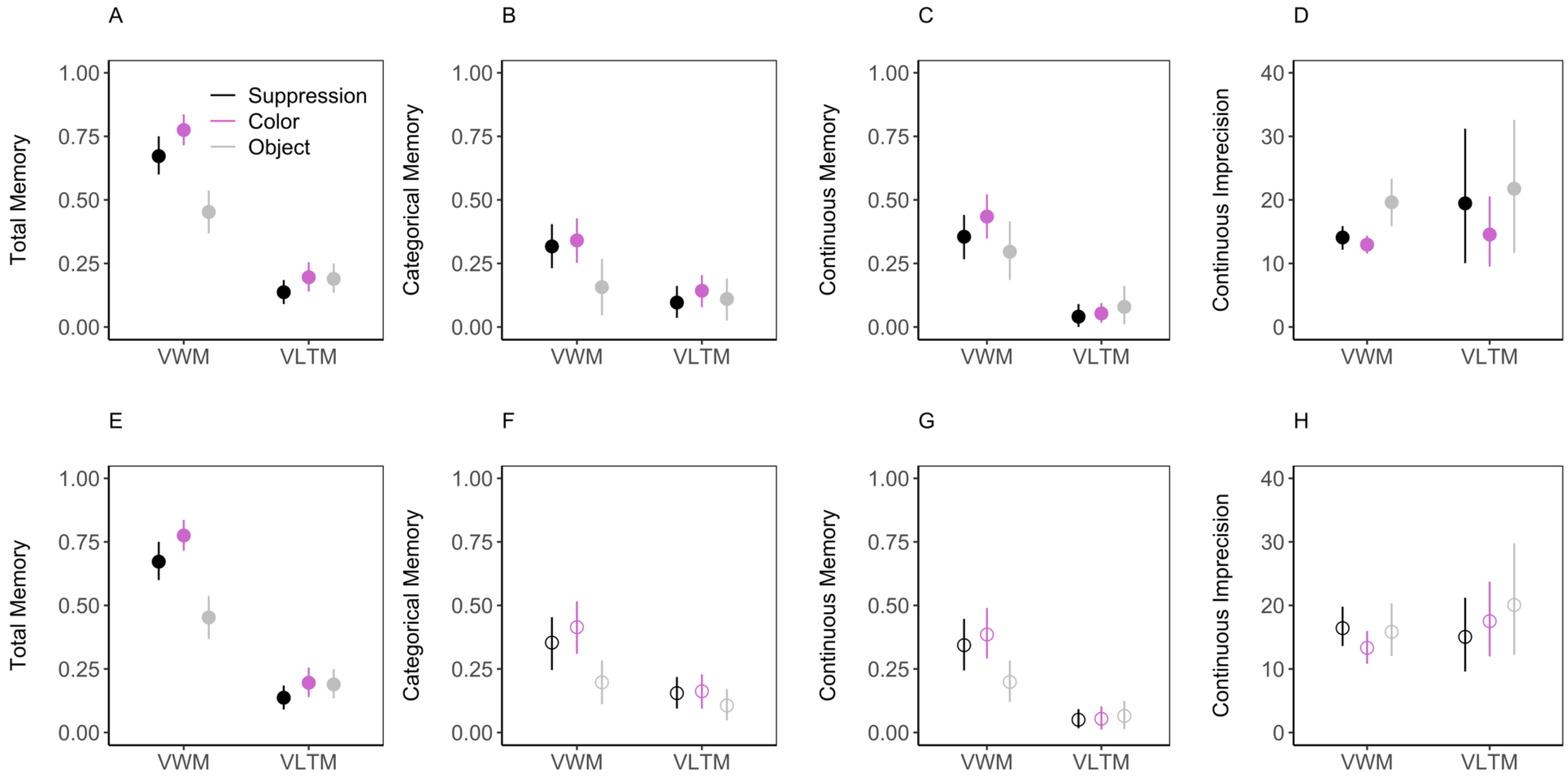
To investigate how labeling affected the storage of categorical and continuous information in visual memory, we submitted our data to mixture modeling. We modeled the data of all participants and conditions simultaneously. We allowed the three main parameters in the model (namely P^M, P^O, and σ^O) to be affected by the two within-subjects predictors of labeling condition (suppression, color labeling, or object labeling) and memory test (VWM vs. VLTM). Each model was restrained to a maximum of seven color categories, with their means taken from the verbal outputs (as shown in Figure 2B). For every model, we ran 10,000 iterations of which the first 1,000 were regarded as burn-in, leaving a total of 9,000 post burn-in iterations for analysis. Appendix A shows that the posterior estimates of all models across all experiments reproduced the actual data.

An aim of this study was to analyse how labeling would change categorical and or continuous information in memory. For this, we then calculated the amount of categorical and continuous information held in memory (categorical = P^M x (1-P^O); continuous = P^M x P^O). Figure 5 presents the mean group-level parameters (dots) and the 95% highest density interval (HDI; error-bars), obtained from the models in Experiments 1a and 1b. These values are also summarized in Table 3. These posteriors should be interpreted as follows: The mean represents the highest point of the posterior distribution and the HDI represents the range of values covering 95% of the posterior distribution. Hence, the HDI indicates the likely values of the parameter given the data. To estimate an effect for or against a verbal labeling benefit, one needs to compare the posteriors of, for example, the labeling condition against the posterior of the suppression condition. If the HDIs of these conditions do not overlap, it gives evidence for a labeling effect as performance between these two conditions substantially differs.

For VWM in Experiment 1a and 1b, color labeling tended to increase total memory (P^M) in contrast to suppression (Figures 4A and 4E). The effect of labeling on these mixture model parameters was however not fully credible as the HDIs of these conditions still overlap. Color labeling had no credible effect on the probability of retrieving categorical information (see Figures 4B and 4F), but it tended to increase continuous memory (Figures 4C and 4G) and reduce memory imprecision (Figures 4D and 4H) in comparison to suppression, but again the labeling effect was not fully credible. In contrast, object labeling led to a reduction of total memory and on the probability of retrieving categorical information compared to suppression. Object labeling also had costs for continuous memory: In Experiment 1a, this was revealed by a reduction in continuous precision (Figure 4D), whereas in Experiment 1b this translated into a lower probability of retrieving continuous representations (Figure 4G). For VLTM, the HDIs of all conditions clearly overlap across all three parameters in both experiments.

Figure 4

Parameter Estimates of Categorical-Continuous Modeling for Experiments 1a (Panels A-D) and Experiment 1b (Panels E-H)



Note. Dots depict the mean of the posterior distributions and the error bars depict the 95% HDI. The two left columns show estimates of probability of retrieving total information, which is defined by the probability to retrieve categorical information (second column from left), and continuous representations (third from left), and the right columns shows estimates of continuous memory imprecision.

Table 3

Posterior means and highest density intervals (HDI) of the Mixture Model Parameters in all Experiments

	VWM						VLTM					
	Categorical		Continuous		Continuous Imprecision		Categorical		Continuous		Continuous Imprecision	
Exp. + Condition + Repetition (R)	Mean	95 % HDI	Mean	95 % HDI	Mean	95 % HDI	Mean	95 % HDI	Mean	95 % HDI	Mean	95 % HDI
E1a Suppression	0.32	[0.23-0.40]	0.36	[0.27-0.44]	14.08	[12.16-15.88]	0.10	[0.04-0.16]	0.04	[1.53×10^{-6}-0.09]	19.46	[10.04-31.20]
E1a Color	0.34	[0.25-0.43]	0.43	[0.35-0.52]	12.96	[11.55-14.37]	0.14	[0.08-0.20]	0.05	[0.02-0.09]	14.55	[9.51-20.55]
E1a Object	0.16	[0.05-0.27]	0.30	[0.19-0.42]	19.63	[15.84-23.32]	0.11	[0.03-0.19]	0.08	[0.01-0.16]	21.76	[11.66-32.57]
E1b Suppression	0.35	[0.25-0.45]	0.34	[0.24-0.45]	16.42	[13.62-19.79]	0.15	[0.09-0.22]	0.05	[0.02-0.09]	15.05	[9.60-21.21]
E1b Color	0.41	[0.31-0.52]	0.39	[0.29-0.49]	13.32	[10.83-15.97]	0.16	[0.09-0.23]	0.05	[0.01-0.10]	17.52	[11.98-23.72]
E1b Object	0.20	[0.11-0.28]	0.20	[0.12-0.28]	15.84	[12.05-20.33]	0.11	[0.05-0.17]	0.07	[0.01-0.12]	20.07	[12.23-29.78]
E2 Suppression R1	0.18	[0.12-0.24]	0.54	[0.46-0.61]	17.19	[16.01-18.50]						
E2 Color R1	0.28	[0.21-0.35]	0.53	[0.45-0.60]	13.04	[11.98-14.03]						
E2 Suppression R2	0.29	[0.21-0.37]	0.65	[0.56-0.72]	13.09	[12.13-14.21]						
E2 Color R2	0.29	[0.21-0.35]	0.68	[0.59-0.74]	11.66	[10.80-12.56]						
E2 Suppression R3	0.22	[0.16-0.29]	0.74	[0.67-0.80]	12.62	[11.78-13.49]						
E2 Color R3	0.27	[0.20-0.34]	0.40	[0.63-0.77]	11.31	[10.46-12.13]						
E2 Suppression	0.28	[0.23-0.34]	0.60	[0.54-0.65]	13.14	[12.43-13.83]	0.28	[0.20-0.36]	0.43	[0.36-0.51]	15.10	[13.78-16.54]
E2 Color	0.29	[0.24-0.35]	0.62	[0.57-0.68]	11.57	[10.92-12.21]	0.38	[0.30-0.46]	0.35	[0.28-0.43]	13.40	[11.82-14.84]
E3 Suppression	0.41	[0.36-0.46]	0.37	[0.31-0.41]	13.82	[12.81-14.72]	0.21	[0.17-0.26]	0.06	[0.03-0.09]	15.22	[11.63-19.03]
E3 Color + Object	0.45	[0.40-0.51]	0.47	[0.41-0.52]	13.84	[12.99-14.68]	0.32	[0.26-0.37]	0.07	[0.05-0.10]	12.40	[9.86-15.24]

9.3.3. Discussion

In both experiments, labeling the color of the colored objects was helpful for the retention of this feature in VWM compared to a condition in which labeling was inhibited with articulatory suppression – as revealed by the recall error measure. With regards to mixture modeling, color labeling tended to increase the accessibility of representations overall and tended to improve memory precision, but in this series of experiments these effects were not credible. These results are in line with the ones of Souza and Skóra (2017) in which color labeling was found to aid the maintenance of color representations in VWM, extending it to a paradigm in which participants maintained color-object bindings. Furthermore, Experiment 1 showed that labeling another feature of the object (its shape) was detrimental to the retention of color information in VWM, reducing both categorical and continuous information. This happened although object-labeling provided a unique cue to the studied object (given that each object was only presented once). These findings rule out several hypotheses of the labeling effect for VWM (see Table 1), namely all hypotheses but hypotheses (4) and (5): labeling the colors seems to activate categorical representations that boost memory for color, whereas labeling other features directs attention away from this feature yielding a cost. Altogether our findings indicate that labeling is only beneficial for VWM if it provides categorical information about the relevant feature of the object.

Critical to our main research question, the delayed test showed that the VWM effects of labeling were short-lived. In line with the results of Kelly and Heit (2017), labeling did not affect VLTM irrespective of whether participants were aware

(Experiment 1b) or not (Experiment 1a) of the upcoming VLTM test. This suggests that labeling impacts visual representations differently over the short- and long-term.

There is one caveat, though. Overall performance in the VLTM test was around 75°. Given that chance performance is this task is associated with a recall error close to 90°, the lack of a labeling effect might be related to poor VLTM learning overall. Simple knowledge about the upcoming VLTM test was not sufficient to yield better performance in this task, given that VLTM performance was similar across the experimental versions in which the delayed test was a surprise (Experiment 1a) or was announced at the beginning of the study (Experiment 1b). It is possible that labeling does foster learning in VLTM, but the number of objects learned (315 in total) and the slim opportunities to commit this information to memory (single study opportunity) precluded us from observing this beneficial effect. The goal of Experiment 2 was to address this possibility.

9.4. Experiment 2

In Experiment 1 participants had a single opportunity to study a color-object binding, and they were only tested on this binding once in VWM and once in VLTM. The objects were studied in conditions that differed in the opportunity to label certain binding features: Participants labeled the color, the object, or they repeated "bababa" aloud to prevent labeling. Color labeling was beneficial and object labeling was detrimental for the storage of the color-object bindings in VWM compared to suppression. Notwithstanding, all conditions yielded the same level of VLTM performance.

These results point to a dissociation between learning over the short- and long-term. Conditions that fostered and hampered VWM had no impact on the retention of representations in VLTM. This is not in line with studies suggesting a link between memory over the short and long-term (Biderman et al., 2019; Brady, Konkle, Gill, et al., 2013; Oberauer et al., 2017). Following up on this issue, Experiment 2 addressed the possibility that labeling did not affect VLTM due to the limited opportunities to learn the color-object associations. Previous studies have shown that long-term learning is fostered by repeated testing of memory compared to restudying (Roediger & Karpicke, 2006; Roediger & Butler, 2011; Roediger & Pyc, 2012). Recently, this *testing effect* was also found to occur for VLTM (Sutterer & Awh, 2016). Sutterer and Awh (2016) presented participants with colored objects for study for a total of 400 images. For half of the images, participants restudied the color a second time; whereas for the other half, they practiced recalling their color. In a final test, participants reproduced the colors of all objects (VLTM test). VLTM performance for the tested objects was higher than for restudied objects. Along with the testing effect, it has been shown that repeated presentation of information also increases VWM performance (Couture & Tremblay, 2006). This repetition effect, also known as the *Hebb effect*, consists of the observation of better recall for memory lists as a direct function of the number of times the list was repeated during the course of the experiment.

The aim of the present experiment was, first, to leverage the repetition and testing effects to increase VLTM performance in the delayed test at the end of the study. Our second aim was to assess whether color labeling could foster long-term learning as reflected in the rate of learning over repetitions (e.g., during the VWM phase). To test for this, the color-object associations were repeated three times in a row (e.g., over three successive VWM trials). Our two main questions were: (1) whether the VWM improvement over repetitions (e.g., the learning rate) would be different across the labeling conditions, and (2) whether this would translate into different performance levels in the delayed recall test in the final VLTM phase.

We predicted that VWM performance would increase across repetitions along with the creation of stronger VLTM traces. Regarding the effects of labeling, we hoped to distinguish between two possible scenarios. One possibility is that labeling only helps over the short term as suggested in Experiment 1. If this is the case, we should observe a labeling benefit in VWM, but labeling should not (a) alter the rate of VLTM learning over the repetitions and (b) it should not yield better recall in the delayed test. Another possibility is that with more opportunities to learn the color-object bindings, labeling would be beneficial both over the short and long-term (e.g., with more learning over repetitions and better delayed recall). This would indicate that the long-term beneficial effect of labeling may be too weak to be observed in single-trial learning but does accumulate over repetitions.

These hypotheses, the experimental design, and the analysis plan for Experiment 2 were preregistered and can be found at: https://osf.io/tker5/.

To foreshadow our results, the color labeling benefit was yet constrained to VWM. We only found a beneficial effect of labeling on the very first exposure to the color-object binding. Over the course of the repetitions, the color labeling advantage vanished within VWM, and it was absent in the final VLTM test. Together with Experiment 1, these results point towards a dissociation on the impact of verbal labeling for memory over the short- and long-term.

9.4.1. Methods

9.4.1.1. Participants

In total, 60 participants (*M* = 23.38, *SD* = 3.89, 42 women) of the University of Zurich took part in this experiment, 58 of these participants had not taken part in an experiment reported here. Participants fulfilled the same criteria and were exposed to same protocol as in Experiment 1a and 1b. Note that we started the experiment with a sample of 30 participants, however as we obtained ambiguous evidence for the interaction of labeling and memory system (VWM vs. VLTM), we increased our sample size until the maximum preregistered sample size was reached.

9.4.1.2. Materials

The same materials as in Experiment 1 were used. In total, 102 objects were chosen randomly for every participant out of the set of 315 objects used in Experiment 1. The color of the objects was randomly assigned and sampled from the same color wheel as in Experiment 1.

9.4.1.3. Procedure

VWM phase. The VWM phase of Experiment 2 followed the same procedure as in Experiment 1b with the following changes. First, Experiment 2 included only two conditions: color labeling and suppression; the object labeling condition was removed. The reason for this was that we wanted to focus on conditions that could improve memory. Second, each VWM trial was presented three times in a row. More specifically, a trial consisted of the sequential presentation of three color-object bindings, and in Experiment 2 the exact same color-object bindings were repeated over three consecutive trials. We thereby lowered the number of objects participants had to learn in contrast to

Experiment 1. Third, the order of presentation of the colored objects varied for every trial repetition to ensure that participants learned the color-object binding (e.g., pink-mug; blue-shoe; green-bucket) and not the order of the colors (pink-blue-green). After every trial, a test phase followed where memory for the colors of the three objects was tested in random order. To simplify, these three trial repetitions are hereafter referred to as one mini-block.

The experiment was divided into six blocks consisting of five mini-blocks each (three with each labeling condition). The manipulation of color labeling and suppression occurred across blocks, which alternated throughout the experiment. Presentation order of the blocks was counterbalanced across participants. In total, participants completed 90 experimental trials consisting of three repetitions of 30 unique sets of three memory objects. Participants learned 90 objects, 45 objects in the color labeling condition and 45 objects in the suppression condition. To familiarize participants with the task, they performed two practice mini-blocks (six trials) of each labeling condition before the exposition to the first experimental block with that condition. The practice blocks were excluded from further analysis. As in Experiment 1b, participants were informed prior to the start of the experiment that they should aim to retain the objects for a longer duration and that they would be asked to recall them again at a second stage in the experiment.

VLTM phase. After the end of the VWM phase, participants completed a multiplication verification task for about 2 min. In this task, simple multiplications (e.g., $3 \times 8 = 25$?) were presented on screen, and participants indicated whether the result was correct or not by pressing the right-arrow key or the left-arrow key, respectively. In total, 40 multiplications were verified. The reason for imposing this task was to eliminate the effect of recency of presentation of the last VWM trials. Next, participants were tested on the colors of the 90 objects learned in the VWM phase in random order. The test was as described for the VWM phase.

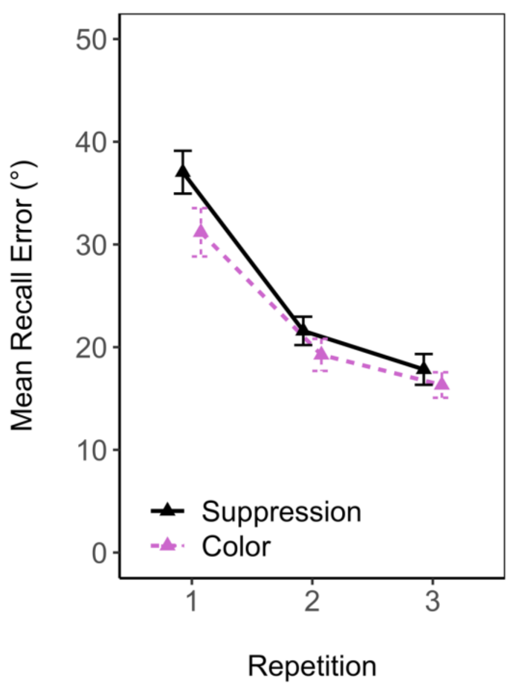
9.4.2. Results

9.4.2.1. Learning Effect on Recall Error

We first assessed the effect of labeling on learning over the three repetitions in the VWM task. Figure 5 shows the mean recall error across repetitions. A color labeling benefit is visible only in the very first exposure to the color-object binding.

Figure 5

Recall Error as a Function of Repetition and Labeling Condition in Experiment 2



Note. Error bars represent the 95% within-subjects confidence interval.

Table 4 shows the analysis of the VWM test including the predictors of labeling condition (suppression vs. color labeling) and repetition (1 vs. 2 vs. 3). The best model of the data included all main effects and their interaction, however there was ambiguous evidence for the inclusion of the interaction as a predictor in the best model even after collecting data of 60 participants [9]. We then followed up analyzing the effect of the interaction by conducting Bayesian ANOVAs between the labeling conditions for each repetition independently [10] (see Table 4). The comparison between the first and second repetition revealed ambiguous evidence against the inclusion of an interaction in contrast to the model with the two main effects. The best model for the comparison between the first and third repetition included the two main effects, but the exclusion of interaction term was again ambiguous. The comparison of the second to the third repetition included both main effects and this model was substantially favored over the model including the interaction between the two predictors.

[9] The interaction was similarly ambiguous for the sample size of 30 participants.

[10] This set of analysis was not preregistered as it was a follow up on the interaction.

Table 4

*Relative Likelihood of Models with Different Fixed Effects Over the Null (BF_{10}) and Relative Likelihood of the Best Model (*e.g., *the One with Higher Likelihood Over the Null) Over the Alternative Model Specified in Each Row (BF_{Best}/BF_{Mrow}) for the Recall Error in the VWM Phase of Experiment*

Rep. (R)	Model n°	Included Fixed Effects: Labeling condition	Repetition	Labeling x Repetition	BF_{10}	BF_{Best}/BF_{Mrow}
All R	**1**	✓	✓	✓	**3.45×10^{55}**	**1**
	2	✓	✓	---	2.86×10^{55}	1.20
	3	✓	---	---	6.61	5.21×10^{54}
	4	---	✓	---	1.67×10^{52}	2.03×10^{3}
R: 1 x 2	1	✓	✓	✓	4.01×10^{28}	1.75
	2	✓	✓	---	**4.61×10^{28}**	**1**
	3	✓	---	---	7.37	6.25×10^{27}
	4	---	✓	---	9.20×10^{25}	501
R: 1 x 3	**1**	✓	✓	✓	**1.24×10^{38}**	**1**
	2	✓	✓	---	6.32×10^{37}	1.96
	3	✓	---	---	1.70	7.28×10^{37}
	4	---	✓	---	5.92×10^{35}	209
R: 2 x 3	1	✓	✓	✓	2.66×10^{4}	4.32
	2	✓	✓	---	**1.15×10^{5}**	**1**
	3	✓	---	---	5.14	1.24×10^{4}
	4	---	✓	---	1.30×10^{4}	8.82

Note. ✓ = effect included in the model. R = repetition.

Finally, we conducted Bayesian *t*-tests contrasting the color labeling and suppression conditions for each repetition independently to estimate a potential color labeling benefit. For the very first presentation, there was strong evidence for a color labeling benefit (BF_{10} = 72.14). For the second and third presentations, however, there was no clear evidence for either the presence of absence of a color labeling effect (BF_{10} =

0.98/ BF_{01} = 1.01; BF_{10} = 0.67/ BF_{01} = 1.49).

Overall Recall Error. Mean recall error between labeling conditions and memory tests is presented in Figure 1. For this analysis, VWM performance reflects the average performance over the three repetitions. Performance was better in the VWM test than in the VLTM test. Similar levels of performance were obtained for the color labeling and suppression conditions in both memory tests.

In the preregistration, we stated that we would analyze the data similarly to Experiment 1. The results of the Bayesian ANOVA are presented in Table 2. The best model included both main effects. This model was preferred over the model including an interaction between labeling and memory system. Furthermore, comparison of the best model against the model with only the effect of memory revealed ambiguous evidence for the inclusion of labeling condition as a predictor. We then followed up analyzing the labeling effect by conducting Bayesian *t*-tests between color labeling and suppression for VWM and VLTM test separately. There was a clear labeling benefit for VWM, BF_{10} = 45.26. In contrast, there was evidence for the absence of a labeling benefit in VLTM, BF_{10} = 0.19 (BF_{01} = 5.26).

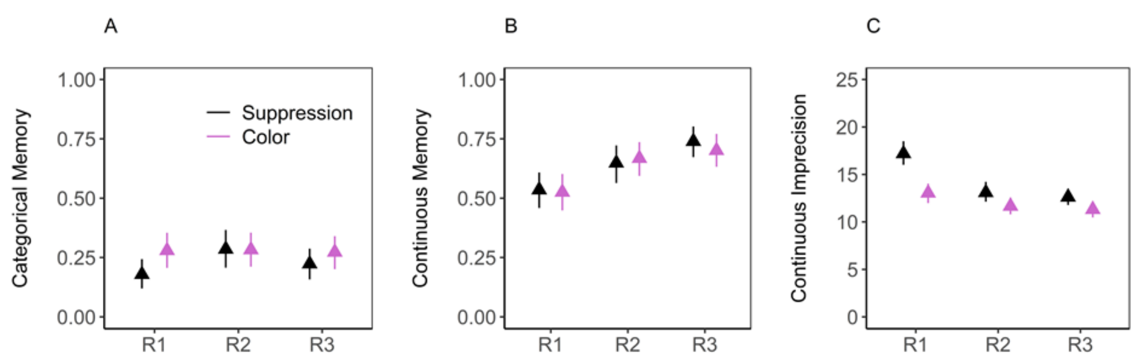
9.4.2.2. Learning Effect on Categorical-Continuous Mixture Model Parameters

In the first model, we assessed the impact of labeling and repeated presentation on VWM. The model included the factor labeling condition (suppression vs. color labeling) and repetition (1 vs. 2 vs. 3). We fitted the model with 10,0000 iterations from which we discarded 1,000 iterations as burn-in, resulting in 9,000 post burn-in iterations for analysis [11].

This model's posterior means and HDIs for our condition of interest can be found in Figure 6 and their respective values in Table 3. The probability of retrieving categorical representations (Figure 6A) was somewhat higher for the color labeling condition compared to suppression condition in the very first presentation of the objects, but this effect vanished across repetitions. The probability of retrieving continuous representations (Figure 6B) was generally not affected by labeling. Lastly, Figure 6C clearly shows that labeling led to more precise continuous memory on the first repetition in contrast to suppression. This boost in continuous precision, however, was substantially reduced in the subsequent repetitions, and it was no longer fully credible.

Figure 6

Mixture Model Parameters (Mean and 95% HDI) for the VWM Data of Experiment 2



Note. R = repetition. Panel A shows probability of retrieving categorical representations, Panel B shows probability of retrieval of continuous representations, and Panel C shows continuous memory imprecision.

[11] We also modeled the data without constraining the color categories. The results of this analyses can be found in the OSF. In general, this analysis yielded a similar pattern to the one reported here.

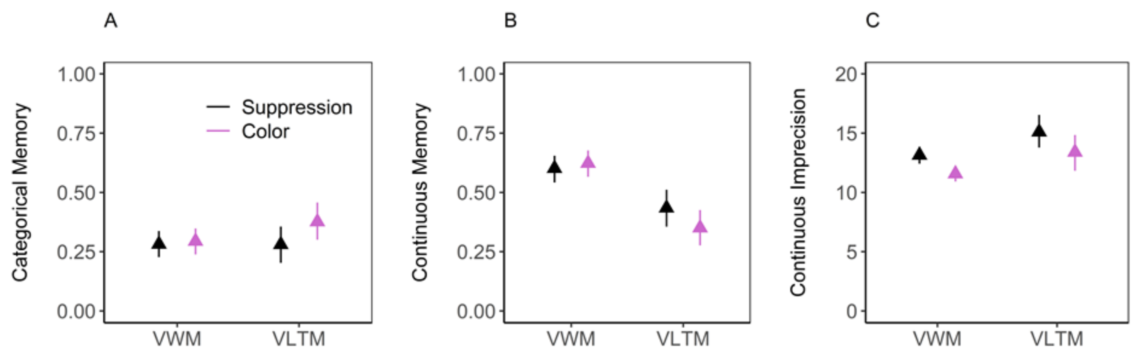
9.4.2.3. VWM vs. VLTM through the Categorical-Continuous Mixture Model Parameters

We then assessed the impact of labeling (color vs. suppression) and the two types of memory tests (VWM vs. VLTM) on the parameters of the categorical and continuous memory mixture model. We again used the color category constraints of the verbal outputs and set the number of categories to seven. The model fit consisted of 10,0000 iterations from which we discarded 1,000 iterations as burn-in.

The posterior means and HDIs for our conditions of interest can be found in Figure 7, whilst the summaries of the estimates are presented in Table 3. Figure 7A shows that categorical memory did not differ between labeling conditions for VWM, but it was somewhat higher for color labeling in VLTM. Figure 7B shows that continuous memory was again not affected in VWM, but for VLTM it was somewhat reduced (although not credibly) by color labeling. Lastly, continuous imprecision was smaller for color labeling than suppression in VWM, and there was a small tendency that this was also the case for VLTM.

Figure 7

Mixture Model Parameters (Mean and 95% HDI) for the Data of Experiment 2



Note. Panel A shows probability of retrieving categorical representations, Panel B shows probability of retrieval of continuous representations, and Panel C shows continuous memory imprecision.

9.4.3. Discussion

In line with Experiment 1, we found a facilitative effect of verbal labeling that was restricted only to VWM despite our efforts to improve long-term learning. Again, the pattern of VWM benefits we observed was in line with the activation of categorical VLTM hypothesis: labeling improved memory precision.

In Experiment 2, we repeated the presentation of the memoranda three times, and this improved performance overall over the short and long-term. Critically, however, labeling the colors did not facilitate learning: improvements over the repetitions were not influenced by labeling and neither was performance in the final delayed test. This addresses one concern raised in Experiment 1, namely, that the beneficial effect of labeling was not detected due to low long-term learning. So far, our results show that labeling the colors of visual objects boosts VWM, but not VLTM. In our last experiment, we assessed whether this result generalizes to conditions in which both the color and the object features are labeled concurrently.

9.5. Experiment 3

The previous experiments all implemented labeling conditions where either the color or the object was labeled, but not both simultaneously. In Experiment 1, labeling the color was beneficial, whereas labeling the object was detrimental to VWM. This raises the question whether labeling both features would yield any benefit at all. The main aim of Experiment 3 therefore was to assess whether labeling the association between the color and the object could be beneficial over the short and the long-term. With regards to VWM, there are three different possible scenarios: (1) The beneficial effect of color labeling is also observed when, in addition to color, the object is labeled; (2) Since labeling the color is beneficial, but labeling the object is costly, these two effects cancel each other out and no effect is observed when both the color and the object are labeled; (3) The impairment of object labeling in VWM prevails when labeling both the object and color. We again tested whether the effects observed over the short-term would be retained when memory is tested after a delay (VLTM test). These hypotheses, the experimental design, and the analysis plan were preregistered and can be found at: https://osf.io/k3nsc/.

To foreshadow our results, labeling the color-object association was beneficial in VWM and, for the first time, we found evidence that this benefit remained in VLTM. This indicates that labeling in VWM only translates into better VLTM when the binding, in this case both the object and its color, are labeled concurrently.

9.5.1. Methods

9.5.1.1. Participants

In total, 60 new participants (M = 24.47, SD = 4.30, 45 women) of the University of Zurich were tested under the same constraints as in Experiment 2. Data of two participants were excluded as they did not comply with the labeling instructions (one did not label at all, and one labeled only the colors on more than 70% of the occasions). We again note that, in line with our preregistration, we first tested 30 participants. As evidence for the effect of labeling across memory systems was in the ambiguous range, we doubled the sample size following our registered plan.

9.5.1.2. Materials

In total, 312 objects were presented to every participant. Colors were assigned randomly to each of the objects.

9.5.1.3. Procedure

VWM phase. The VWM phase of Experiment 3 followed the same procedure as in Experiment 1 with the following exceptions: First, Experiment 3 included two labeling conditions: color + object labeling vs. suppression. In the color + object labeling condition, participants were instructed to overtly label the presented color and the object (e.g. "blue heart"), whereas in the suppression condition participants were instructed to articulate “bababa” aloud. Second, in this experiment every trial consisted of the sequential presentation of three objects, with each object being onscreen for 250 ms, followed by a 2250 ms inter-stimulus blank interval. The inter-stimulus interval was increased to accommodate for the fact that labeling the binding takes longer than labeling only one single aspect of the stimulus. Accordingly, more time was provided for the suppression condition. The color + object labeling and suppression trials alternated every 10 trials throughout the experiment, and the order of labeling conditions was counterbalanced across participants. The experiment consisted of 104 trials, 52 for each labeling condition, of which the first two trials in each block were regarded as practice trials, resulting in 50 experimental trials in each condition. As in Experiment 1b, participants were informed that they needed to recall the objects at a later point in time and were asked to try to remember them for a longer period.

VLTM phase. After the VWM task participants took a short break, in which they left the experimental room and were offered some sweets (e.g. chocolate). Then, participants were tested again on the colors of the 300 objects (12 objects from the practice trials not included) learned in the VWM phase in random order.

9.5.2. Results

9.5.2.1. Recall Error

The mean recall error for each memory test and labeling condition are visualized in Figure 1. Visual inspection clearly shows that performance in VWM is better than for VLTM, in line with all of the previous experiments. There is a benefit for labeling the color+object association in VWM compared to saying "bababa". For the first time in our series of experiments, there was a labeling benefit in VLTM, as the recall error in the color+object labeling condition was smaller than in the suppression condition.

We preregistered to analyze the data in accordance with the previous experiments. The results of the Bayesian ANOVA are presented in Table 2. The best model of the data included the main effects of labeling and memory test. However, there was ambiguous evidence for excluding the interaction of labeling and memory test from the best model, even after we collected data of 60 participants. Bayesian *t*-tests yielded evidence for a clear labeling effect within VWM, $BF_{10} = 2.30 \times 10^{13}$, and also within VLTM, $BF_{10} = 1.98 \times 10^{7}$. Hence, the ambiguous interaction is not due to labeling not being beneficial over the long-term, but it seems to relate to ambiguous evidence regarding whether this benefit is of the same size in VWM and VLTM. Regardless of whether this benefit is of the same size or not, the critical point is that Experiment 3 showed, for the first time, evidence for a labeling benefit in episodic VLTM. This suggests that a long-lasting labeling benefit is constrained to conditions in which bindings are labeled.

9.5.2.2. Categorical-Continuous Mixture Model

As in the previous two experiments, we assessed the impact of labeling and the two memory tests on categorical and continuous memory along with continuous imprecision. We again used the color category constraints of the verbal outputs and set the number of categories to seven. The model included the factor labeling condition (suppression vs. color+object labeling) and memory (VWM vs. VLTM). The model fit consisted of 10,0000 iterations from which we discarded 1,000 iterations as burn-in.

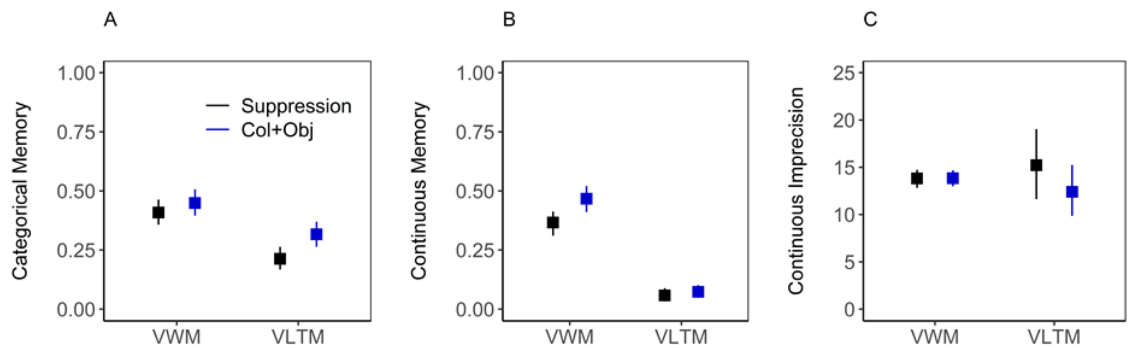
The posterior means and HDIs for our conditions of interest can be found in Figure 8 and the summary of the estimates in Table 3. Figure 8A shows that categorical memory was somewhat higher in the color+object labeling condition in comparison to suppression in VWM. The same pattern is visible for VLTM, but here the increase in categorical memory is clearly credible. Continuous memory (Figure 8B), in contrast, was only higher for the labeling than the suppression condition for VWM, but not for VLTM. Lastly, continuous imprecision (Figure 8C) did not show a labeling benefit, neither in

VWM nor in VLTM. For VLTM, there is a small but not credible tendency of a decrease due to labeling.

We also fitted the model allowing free estimates of the color categories and the model was fairly in line with the reported model with the constrained color categories, except that for continuous memory, the labeling benefit was less high and not credible.

Figure 8

Estimated Mixture Model Parameters (Mean and 95% HDI) for the Data of Experiment 3



Note. Panel A shows the probability of retrieval of categorical representations, Panel B shows the probability of retrieval of continuous representations, and Panel C shows continuous memory imprecision.

9.5.3. Discussion

Experiment 3 replicated the finding of a labeling benefit in VWM in contrast to a suppression condition. This time, labeling was not only helpful when the color itself was labeled, but rather when the color and object binding was labeled extending the scope of the labeling effect in VWM. This stands in contrast to the fact that when only the object was labeled in Experiment 1, it led to the forgetting of the color (as if it led to the filtering of this information). However, when the object was labeled alongside the color, it no longer competed with the relevant color information, and both features could be stored.

For the first time in our series of experiments, we could show that a labeling benefit in VWM was translated into better VLTM. When analyzed with the mixture model, the data showed that the sources of the labeling benefit were different between VWM and VLTM: Replicating our previous experiments and Souza and Skóra (2017), labeling improved storage of continuous representations in VWM. In contrast, for VLTM, the beneficial effect of labeling was mainly due to categorical representations, with no credible changes to continuous memory.

9.6. General Discussion

Across three experiments, we found a labeling benefit in VWM when participants labeled the color of a colored object. In Experiment 1 and 2 we showed that color labeling benefited continuous color recall in VWM compared to a suppression condition. Additionally, Experiment 1 showed that labeling the object's identity yielded a cost to the retention of object's color compared to suppression. This indicates that color information is lost when participants label another feature of the visual object (e.g., its shape). When both the color and object's identity were labelled concurrently though (Experiment 3), the labeling benefit in VWM also emerged in comparison to suppression. These findings confirm that labeling affects the storage of visual information in VWM, extending the results reported by Souza and Skóra (2017). They show that labeling adds information to the visual features stored in VWM, and this can lead to augmented retention of the labeled feature, even if this may come at the expense of the non-labeled features. In the particular case of Experiment 1, labeling the object identity seems to have led to the filtering of the color information.

In contrast to a labeling benefit in VWM, we could not find evidence for a labeling benefit for the retention of the same objects for a delayed recall (VLTM) test in Experiments 1 and 2. This was the case when participants were not aware (Experiment 1a) and aware of the VLTM test (Experiment 1b) prior to the start of the experiment. Moreover, in Experiment 2, we ruled out the possibility that this lack of effect was due to rather poor VLTM in general. In Experiment 2, participants repeatedly saw the same color-object pairs for three consecutive trials, thereby fostering learning in VLTM by means of the repetition (Couture & Tremblay, 2006; Johnson et al., 2017; Lafond et al., 2010; Oberauer & Meyer, 2009) and the testing-effect (Roediger & Butler, 2011; Roediger & Karpicke, 2006; Roediger & Pyc, 2012; Sutterer & Awh, 2016). This manipulation substantially improved delayed recall, yet no labeling benefit was observed for VLTM. These findings are in line with the lack of a color labeling benefit observed by Kelly and Heit (2017).

These experiments suggest that the beneficial effect of labeling on VWM observed by Souza and Skóra (2017) and the lack of a labeling benefit on VLTM observed by Kelly and Heit (2017) are not due to differences in the procedures used to induce labeling (aloud responses vs. keypress) and to test memory (continuous color reproduction vs. color hue recognition test). Here, we maintained these features constant and were able to show the same dissociation in the retention of labeled information over the short- and long-term. The only experiment in which we could obtain a labeling benefit both in VWM and VLTM was Experiment 3, wherein the color and object identity was labeled together.

9.6.1. Implication of Verbal Labeling for Continuous and Categorical Representations in Memory

One aim of the present study was to analyze the contribution of verbal labeling to the storage of coarse (categorical) and more fine-grained (continuous) visual representations over the short- and long-term. This was assessed by modeling the data with a mixture model that attempts to distinguish between the sources of information used to respond in the task, namely categorical information about the colors, continuous information about the precise hue studied (and the precision of this information), or guessing.

For VWM, mixture modeling of all experiments indicated that labeling the color of an object increased the probability of retrieval of this information overall as opposed to guessing, replicating Souza and Skóra (2017). These authors further showed that this benefit was not due solely to addition of categorical representations: either the probability of continuous information in memory increased while continuous precision remained relatively the same; or continuous precision increased along with little change in the amount of continuous information stored. In the present study, we found a similar mix of effects: the quantity of continuous memory increased in Experiment 3, whereas we found rather improvements in memory precision in Experiments 1 and 2. This means that labeling allowed detailed information from a larger number of items to be stored (e.g., effect on continuous memory parameter), or that the number of items for which continuous information was retained remained the same, but their continuous recall was more precise (e.g., effect on continuous imprecision parameter).

To the best of our knowledge, the categorical-continuous mixture modeling approach has not yet been used to assess VLTM nor the role of labeling therein. Our experiments showed that, in general, information stored in VLTM had a lower probability of retrieval and lower precision compared to VWM, replicating prior findings (Biderman et al., 2019). The lower VLTM precision was observed although the model controls for categorical responding, which in itself would be associated with lower precision in mixture models that do not include categorical responses. This shows that the lower precision of VLTM representations cannot be accounted by larger proportion of categorical responses in delayed tests. Furthermore, in Experiments 1a, 1b and 3, we also observed that the probability of retrieving categorical representations was higher than of retrieving continuous representations in VLTM, whereas for VWM the division between categorical and continuous representations was more even. This suggests that another differentiating factor between VLTM and VWM may pertain to the retention of continuous information.

Regarding the effects of labeling, there was no labeling effect on VTLM when the data was modeled in Experiments 1a and 1b in agreement with the results obtained for recall error. In Experiment 2, labeling tended to increase categorical information at the expense of more continuous information (a small reduction on probability of retrieval and on precision), such that average performance did not improve (as indicated by the recall error data). In Experiment 3, labeling improved performance as revealed by the recall error analysis, but again mixture modeling indicated that this benefit was associated with increases in categorical memory only (with continuous memory remaining unchanged), unlike what was observed for VWM. Hence, labeling of both features seems to play an important role for a labeling benefit that can be maintained across a longer time-period into VLTM. The novel insight provided by this experiment was that the labeling benefit in VLTM reflected an increase in categorical information with no change in continuous memory, again in stark contrast to the effects observed for VWM.

9.6.2. Different Role of Labels in VWM and VLTM

In the introduction we discussed five hypotheses of the effect of verbal labeling in visual memory. Our results help distinguishing between the plausibility of these hypotheses as likely explanations of the labeling effect in VWM and VLTM.

First, our results do not support the *verbal recording* hypothesis, neither for VWM nor VLTM: across most experiments, we did not find an indication that labeling increased categorical representations at the expense of continuous information or its precision as predicted by this hypothesis. For VLTM, some studies have found a cost for labeling in line with the verbal overshadowing effect (Brandimonte et al., 1997; Lupyan, 2008; Schooler & Engstler-Schooler, 1990). The only instance in which we observed a tendency for a trade-off between categorical and continuous information in VLTM was in Experiment 2. This trend was not credible though.

Second, the *dual-trace* hypothesis predicts that labeling would only increase categorical responding with no change in continuous memory. This prediction fits with the labeling benefit observed for VLTM in Experiment 3. This hypothesis, however, cannot explain the VWM data.

Third, the *distinctiveness* hypothesis predicts that the labeling benefits would be proportional to how much the label differentiates between the memoranda. In Experiments 1a and 1b, we included an object labeling condition that allowed the generation of a unique label for each item in the experiment (since each object was only presented once). Contradicting this hypothesis, this condition yielded costs to VWM performance, and no effect for VLTM retrieval.

Fourth, the *activation of categorical VLTM* hypothesis predicts that labels activate VLTM representations of the category. This would allow people to store more visual details because the individual item's properties can be stored in relation to the category. This may facilitate data compression or the use of hierarchical representations that reduce memory load (Brady et al., 2009). In line with this hypothesis, VWM performance benefited from color labeling by showing an increase in continuous memory or continuous precision (see also Souza & Skóra, 2017). This effect however was constrained to VWM; VLTM did not show increases in continuous memory as a function of labeling.

Fifth, the *cue to focus attention* hypothesis predicts that labeling guides attention to the labeled features, and this can be helpful or harmful depending on the match between the attended feature and the relevant feature. In our experiments, participants were fully aware that color was the relevant feature, hence color labeling could not be beneficial according to this hypothesis. Object labeling, however, would direct attention away from the relevant feature and hence this hypothesis predicted a cost in this condition. Our data partially matches those predictions: on the one hand, this hypothesis fails to account for the fact that color labeling does improve memory, specially VWM, but also VLTM if color labeling is combined with labeling the object. On the other hand, it correctly predicts a cost for object labeling in VWM. This suggests that labels serve to guide attention to certain features, but this does not fully explain the resulting benefits that follow from it.

To conclude, we found evidence in partial support of three mechanisms: (1) verbal labels guide attention to the labeled feature, and this differential attention affects VWM processing, (2) the label activates categorical knowledge in VLTM, and (3) for VWM, this VLTM activation allows for storage of more visual details, perhaps because categorical information permits exploitation of redundancies in the visual input (e.g., facilitating data compression or creation of hierarchical representations). These more precise representations created in VWM, however, either are (a) not transferred to VLTM or (b) they do not seem to survive the proactive interference accumulated in VLTM as more and more objects are learned. As such, at best, knowledge activation through labels only serves to increase categorical storage in VLTM, and only if this activation is combined with the concomitant activation of the retrieval cue (e.g., the object's label).

9.6.3. Creation of Representations in VWM and VLTM

In order to create a durable memory representation, the visual object needs to be perceived, encoded, and consolidated to be later accessible in memory (Cowan, 2017; Ricker, 2015). Attention and time are assumed to be necessary to create a stable memory representations both in VWM (Ricker & Cowan, 2014; Ricker & Hardman, 2017) and VLTM (Huebner & Gegenfurtner, 2011). During encoding, a visual trace of the memory object is built up, which is then transformed into a memory representation by the process of consolidation (Ricker, 2015; Ricker et al., 2018). So far, it is unclear whether consolidation creates a representation that is accessible both over the short- and the long-term (or whether there are two separate consolidation processes). Labeling might operate by facilitating the consolidation of the memory representation in VWM. Why labeling would have a limited impact in VLTM consolidation, however, is unclear. This may depend on how memory models conceptualize the relation between VWM and VLTM (Cowan, 2008).

There are several models of VWM and VLTM that make different assumption about the relation between these two systems. Some models assume that VWM and VLTM are completely independent systems, with VWM being a limited-capacity memory system and VLTM an unlimited one (Atkinson & Shiffrin, 1968). Thus, these models would predict that representations in VWM and VLTM are created separately and independently. In line with this possibility, Brady, Konkle, Alvarez, et al., (2013) observed that over the short-term, the color (e.g., pink or green) and state (e.g., open or closed) of objects were remembered in conjunction; after a longer delay (and the learning of several objects), color was forgotten much faster than state. This suggests that representations accessible over the short-term may seem maintained as a conjunction, but over the long-term they may be lost independently. Likewise, this assumption of separate VWM and VLTM stores would fit with our finding of a dissociation of the labeling effect between VWM and VLTM: we found evidence for an increase in continuous information in the short-term and an increase in categorical information over the long-term.

The assumption of purely separate systems does not, however, suffice to account for the effects of labeling we have observed. Recent research from our lab has shown that labeling activates categorical knowledge in VLTM, and this activation critically contributes to the boost in memory precision observed in VWM. Souza et al. (2020) showed that the categorical distinctiveness of the labels is directly related to their memory precision benefit: when the labels categorize the memoranda into two broad categories, memory precision decreases. When labels divide the stimuli into more categories (4 or more), precision increases. This shows that categorical knowledge stored in VLTM affects ongoing processing in VWM, changing how information is represented. This implies a tight link between VWM and VLTM, rather than static separate systems. There are several models that view WM as an activated subset of LTM (Cowan, 1988; Oberauer, 2009; Oberauer & Hein, 2012). In these models, WM has been described as consisting of a focus of attention and the activated part of LTM. Oberauer and Hein (2012) assumed that the focus of attention could be divided into a narrow focus, holding only one single object and thereby giving it a special role, and a broad focus of attention, reflecting the focus of attention in Cowan (1988), where about a handful of objects are accessible. Lastly, representations in the focus of attention are assumed to be part of VLTM: they are activated VLTM representations. Within such models, verbal labeling can be conceived as a means to activate VTLM, thereby facilitating the binding of information into the broad and narrow focus of attention. It is also possible that the activation of VLTM representations do not boost VLTM learning due to interference: activation in VLTM spreads quickly, and without a mechanism such as the focus of attention, activation of multiple representations led them to interfere with each other.

Lastly, there is the possibility that the increase of categorical information in VLTM (in Experiment 3) was a confound of our task setup: the VWM test may have promoted an increase in the categorical information at the expense of the continuous. Once participants made a first response about the color of a memory object in the VWM test, it may have given participants the incentive to store a more categorical representation. We have three arguments against this possibility. First, we first did not replicate this effect of a categorical increase in VLTM in Experiment 1, where the exact same experimental procedure was used. Second, in Experiment 2 we repeatedly exposed participants to the VWM test on three occasions. If testing would promote an increase of categorical information, then we should have seen such an increase over the repetitions in this experiment, possibly at the expense of the continuous information. Our results are not in line with this assumption as we saw improvements in memory precision over repetitions and increases in continuous memory. Thus, the VWM test most likely did not promote an increase of categorical information at the expense of continuous information.

9.6.4. Verbal Labeling Benefit in Relation to Retrieval Practice

In Experiment 2 we replicated the beneficial effect of repeated studying and testing on both VWM and VLTM (Roediger & Karpicke, 2006; Roediger & Butler, 2011; Roediger & Pyc, 2012; Sutterer & Awh, 2016). In the VWM task, each color-object pair was presented and tested three times while participants labeled the colors or said “bababa” aloud throughout the repetitions. The labeling benefit was restricted to the very first exposure to the object and vanished for the second and third repetition, contributing further evidence that the verbal labeling effect is short-lived and does not affect the rate of learning.

Relatedly, the absence of a verbal labeling effect for VLTM in general, and with the repeated presentation of the colored objects rules out an explanation of the verbal labeling effect as retrieval practice. One could argue that in order to label, one has to retrieve this information, thereby leading to an additional retrieval practice not present in the suppression condition. This retrieval practice could explain the beneficial effects of labeling observed by Souza and Skóra (2017) and in Experiment 1. If this was the case, we should expect labeling to improve VLTM since we know retrieval practice does improve VLTM retention (Sutterer & Ahw, 2016). Furthermore, performance in the second presentation of the colored object in the labeling condition would imply four retrievals (two in the first trial + two in the second trial), and hence it should have been even better than performance in the third repetition in the suppression condition. Experiment 2 showed, however, that performance improved linearly with the number of repetitions in VWM irrespective of labeling. This is inconsistent with the possibility that labeling benefits VWM through retrieval practice.

9.7. Conclusion

The way in which we describe our visual surroundings can have a profound impact on the visual memories that are formed to guide our behavior over the short- and long-term. Here we demonstrated for the first time that verbal labeling is either beneficial or inconsequential for the retention of visual memories, and that the source of this benefit is different across short and long timescales. Verbal labels provide categorical information that boosts the maintenance of high-fidelity representations in VWM to guide our immediate behavior. These detailed representations are either not retained over the long-term or they do not survive interference that accumulates in VLTM. As such, verbal labeling can, at best, allow for the retention of more categorical information over the long-term.

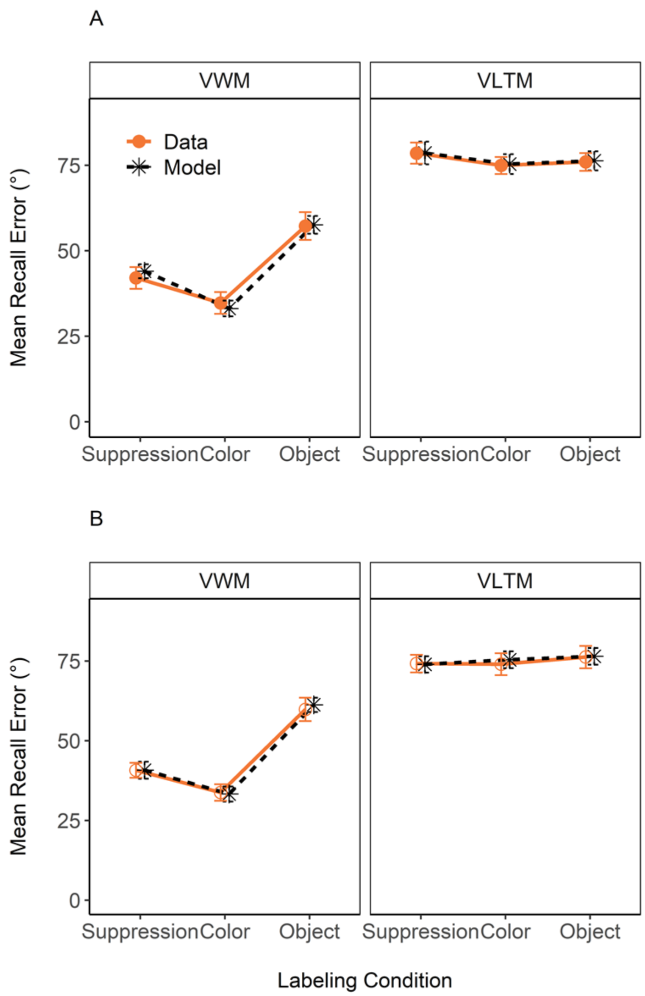
9.8. Appendix A

9.8.1. Model Fit

To assess how well the model captured the data, a posterior predictive check was performed by simulating data (predictions) based on the full model parameters for all experiments. Figure A1A and A1B show that the predicted recall error seemed to be fairly in line with the data for Experiment 1a and 1b, respectively. For Experiment 2, Figure A2A shows that the modeling fit the data for Experiment 2 for the three repetitions in VWM, and Figure A2B for the VWM and VLTM model. Figure A3 shows that the posterior estimates of the model in Experiment 3 also reproduced the actual data.

Figure A1

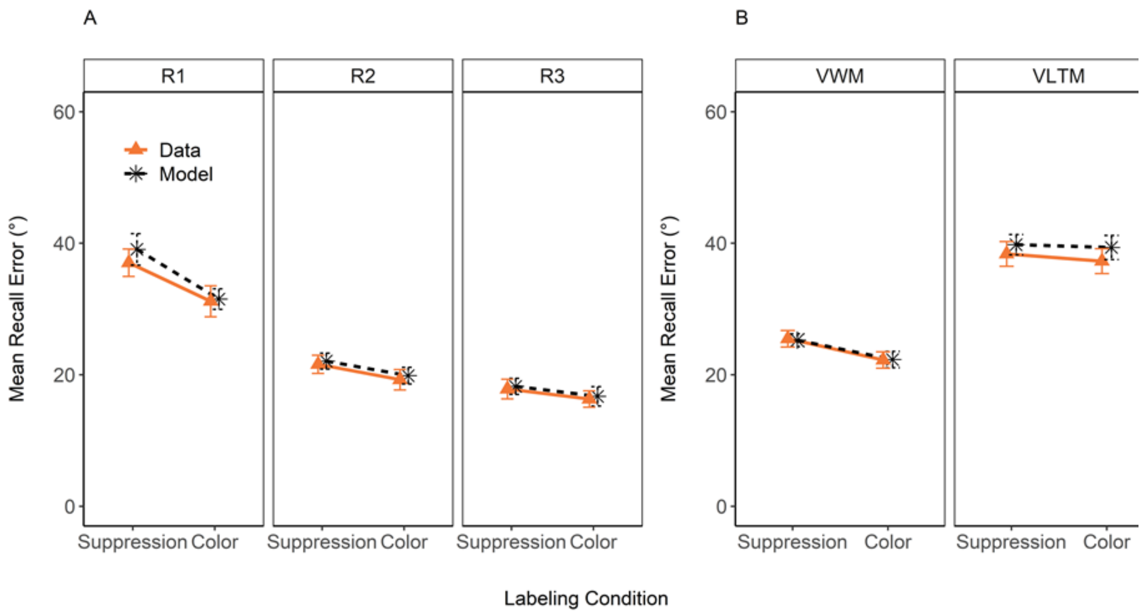
Recall Error Obtained for the Data of Experiments 1a and 1b and the Predicted, Simulated Data from the Posterior Estimates of the Mixture Model Fitted to this Data



Note. Error bars represent the 95% within-subjects confidence interval.

Figure A2

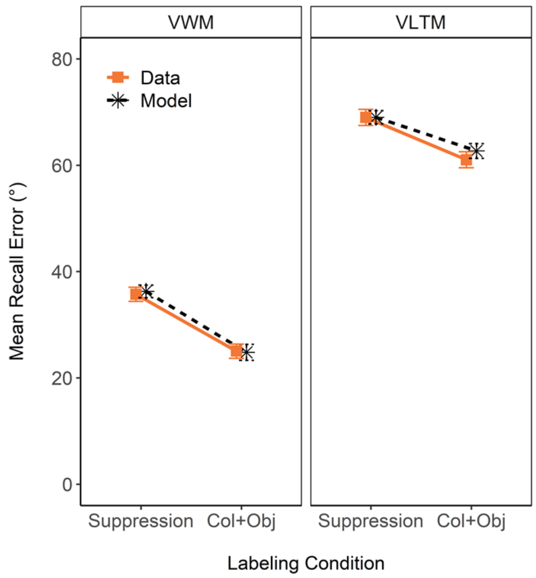
Recall Error Obtained for the Data of Experiment 2 and the Predicted, Simulated Data from the Posterior Estimates of the Mixture Model Fitted to this Data



Note. Panel A show the data and predictions for the trial repetitions (R1, R2, R3) as a function of labeling condition in the VWM phase. Panel B shows data and predictions for the model comparing VWM performance (averaged across repetitions) and VLTM. Error bars represent 95% confidence interval.

Figure A3

Recall Error Obtained for the Data of Experiment 3 and the Predicted, Simulated Data from the Posteriors of the Mixture Model Fitted to this Data



Note. Error-bars represent 95% within-subjects confidence intervals.

10. The Fate of Labeled and Non-Labeled Visual Features in Working Memory

Clara Overkott and Alessandra S. Souza

University of Zurich

Status: *Submitted* to Journal of Experimental Psychology: Learning, Memory and Cognition (07.08.2020)

Author's contribution

Clara Overkott: Development of the specific research questions, programming the experiments, analyzing the data, writing the manuscript

Alessandra Souza: Supervision of the project, development of the general research idea, programming of the experiments, discussing the research questions and results, commenting on the manuscript

Acknowledgments

This research was support by a grant from the Swiss National Science Foundation (100019_169302) awarded to A. S. Souza. We thank Gian-Luca Gubler, Natascha Zimmerli, Vanessa Vallesi, Jasmine Kerr, Olenka Dworakowski and Elena Throm for helping with data collection. The pilot study and one out of the three experiments reported in this paper were pre-registered in the Open Science Framework (OSF). Pre-registrations, materials, data, and analysis scripts can be found on: https://osf.io/z3yp3/.

10.1. Abstract

We often verbally describe or label the visual world around us. Verbal labeling has been found to improve the retention of visual information in visual working memory (VWM). Visual objects, however, often consist of a conjunction of several features. What happens to the memory of this object when we label only one of its features? Here, we used multi-feature objects (varying in color, orientation, and spatial frequency) to assess the fate of labeled and non-labeled features in VWM. While encoding the memoranda, participants labeled either its (a) color, (b) orientation, or (c) spatial frequency; or repeated "bababa" aloud thereby inhibiting verbal labeling. At test, these features were reproduced using a continuous scale. Across three experiments, we found that labeling resulted in an increase in the amount of detailed memory about the labeled features compared to the suppression condition. The impact on the non-labeled feature was more mixed across features: color memory was always impaired when other features were labeled. Orientation memory showed almost no impact of color labeling, but was impaired when spatial frequency was labeled. Spatial frequency, conversely, showed no costs due to color or orientation labeling. In sum, verbal labeling produced asymmetric effects in VWM: detailed memory for the labeled features always received a boost, whereas detailed memory of some of the non-labeled features was involuntarily filtered. These results point to trade-offs in the allocation of VWM capacity as a function of labeling.

10.2. Introduction

Visual working memory (VWM) is the system that holds and maintains visual information available for immediate processing. The total amount of visual information a person can maintain in VWM is severely limited (Oberauer et al., 2016). Verbal labeling can counteract the capacity limitations of VWM: previous research has demonstrated that describing visual objects, for example by saying *"this is a blue building",* can help to retrieve detailed visual information regarding the color of the building a moment later (Overkott & Souza, 2020; Souza & Skóra, 2017; Souza et al., 2020). But, does this come at the expense of remembering other non-labeled features of the labeled object (e.g., its size, material, shape, orientation)? Visual objects usually contain multiple features, and the more features of an object one retains in VWM, the worse the memory for each individual feature becomes (Fougnie et al., 2010; Hardman & Cowan, 2015; Oberauer & Eichenberger, 2013; Olson & Jiang, 2002; Palmer et al., 2015; Quak et al., 2018; Swan et al., 2016). This indicates that there is a common capacity limit for storing different features in VWM. Accordingly, it is possible that boosting the precision of one visual feature in VWM may come at the expense of the retention of other memory features. The main aim of the present study was to investigate whether describing one feature of a visual object can induce a trade-off on the retention of the labeled and the non-labeled features in VWM.

In the following, we will first describe studies investigating the impact of generating descriptions for the retention of visual information in VWM. Then, we will present current hypotheses regarding the impact of verbal labeling on visual memory. Next, we elucidate the reasons why labeling one feature of a multi-feature object may lead to the filtering of the non-labeled feature and finally present our research aims.

10.2.1. What Happens When Visual Information is Labeled?

Recent research has shown that verbal labeling improves VWM (Forsberg et al., 2019, 2020; Overkott & Souza, 2020; Souza et al., 2020; Souza & Skóra, 2017). For example, Souza and Skóra (2017) sequentially presented four colored discs for study and asked participants to either label the colors or repeatedly say “bababa” aloud to prevent labeling (aka suppression condition). During test, the colors of all four discs were reproduced on a color wheel. Their results showed that verbally labeling the color improved memory performance in contrast to suppression. But what was the source of this benefit? In order to answer this question, the data needs to be analyzed through computational models that can separate the differential sources of influences on the responses. Data of this type of task (known as delayed estimation) can be submitted to mixture modeling to estimate the proportion of responses that included information retrieved from VWM (as opposed to guessing) along with the fidelity or precision of the information in memory (Bays et al., 2009; Zhang & Luck, 2008). Recently, these models have been extended to account for categorical biases (Bae et al., 2015; Hardman et al., 2017; Pratte et al., 2017). These models assume that memory responses are based either on (a) categorical information, such as memory that an item belonged to the blue or left category, or (b) continuous information, e.g., the exact hue of blue or exact direction the item pointed to. Mixture modeling in Souza and Skóra (2017) revealed that the labeling benefit accrued from two sources: (1) participants retained categorical information about more items, and (2) they also retained more continuous information: they either had a great probability of remembering the exact color hue of a larger proportion of the memory items or they stored this continuous information more precisely.

This finding has been replicated in different task set-ups. For example, Overkott and Souza (2020) presented three colored objects sequentially and asked participants to label the color (e.g., “blue”), the object (e.g., “dog”), or the color+object binding (e.g., “blue dog”); and again contrasted these conditions to suppression. At test, participants reproduced the color of the object with the use of a color wheel. Labeling the color or the color+object improved performance in contrast to suppression. Mixture modeling revealed that color labeling increased the accessibility of representations (categorical and continuous) and memory precision for the continuous information, whereas color-object labeling increased the probability of storage of continuous information. Critically, in this study, labeling only the object produced costs compared to the suppression condition. This result suggests that labeling can also lead to costs depending on which feature is labeled and which feature needs to be retrieved from VWM.

Another study from our lab further showed that the beneficial effect of labeling depends on the categorical distinctiveness of the labels. Souza et al. (2020) manipulated the number of labels used to categorize continuously varying colors or shapes. In the color task, participants self-selected 2, 4, or any term they wanted to label the colors. Performance in the delayed estimation task monotonically increased with the number of labels used compared to suppression. Mixture modeling showed that the use of very broad category labels (2-labels) decreased memory precision, whereas more specific labels (more than 4 labels) increased memory precision. These effects were replicated in a continuous shape task in which the labels were non-words and categorization of the continuous shape-space was learned in the laboratory.

To summarize, recent research suggests that verbal labeling is especially helpful for the retention of detailed information in VWM: labeling increases the probability of storing continuous information about more items or the precision with which this information is stored. However, labeling may also incur in costs as when the labels are broadly applied to categorize the memoranda. Some initial evidence also points to trade-offs between labeled and non-labeled features: labeling one feature could lead to the loss of the non-labeled feature (Overkott & Souza, 2020). However, investigating these trade-offs were not the main goal of Overkott and Souza (2020), and it is not clear whether this happened because the labeled feature (the object's shape) was less relevant in the task because it only served as a retrieval cue. Investigating this further will be the main goal of the present study.

10.2.2. Hypotheses of the Labeling Benefit

Researchers have commonly assumed that labeling would only provide categorical knowledge about the visual trace (Alogna et al., 2014; Donkin et al., 2015; Lupyan, 2008; Schooler & Engstler-Schooler, 1990; Sense et al., 2017). On the worst-case scenario, the verbal label would overshadow the visual input leading to the loss of the visual trace, and hence to less precise memory (a hypothesis known as *verbal recoding*, see Souza & Skóra, 2017). On the best-case scenario, the label would just add another source of information (i.e., categorical knowledge) with no change on the visual trace (i.e., a *dual-trace* hypothesis). This would predict only an increase in categorical knowledge in the presence of the label with continuous information remaining constant. Both of these hypotheses cannot account to the benefits of labeling observed for the retention of continuous information described above.

To account for these findings, Souza and Skóra (2017) proposed a *categorical visual long-term memory* (VLTM) hypothesis that assumes that verbal labeling activates categorical visual information in VLTM (see also, Lupyan, 2012), and that this categorical knowledge facilitates the retention of continuous information in VWM. This hypothesis assumes that two visual memory traces are created in memory: one consisting of the visual representation of the presented item and the other one of the visual information that was activated in VLTM by the verbal label. As the feature category is activated in VLTM, it is possible that this either facilitates the encoding and consolidation of the visual memory trace in VWM such that it now allocates more precise continuous information regarding the labeled visual feature, or the categorical activation prevents the continuous information from interference during maintenance or retrieval, sustaining it in a more robust state.

There are other hypotheses of the labeling effect, for example, that labeling would act as a cue to focus attention on the labeled feature (Kelly & Heit, 2017). According to this *attentional-cue* hypothesis, the verbal label would be helpful when it directs attention to an item's feature making it relevant. Likewise, the label would induce a cost when attention is directed to irrelevant information, because it detracts attention away from the relevant information. Kelly and Heit (2017) found that labeling the color during study reduced color bias towards a color prototype in a surprise VLTM test compared to study conditions involving animacy judgement or preference rating. The authors argued that this was due to the label directing attention to the color feature during the study phase, which helped performance when this feature suddenly became relevant during the delayed memory test. However, this effect vanished once participants were made aware of the memory test already before the study phase. Although this hypothesis as formulated does not account for the labeling benefits observed by Souza and Skóra (2017), Overkott and Souza (2020), and Souza et al. (2020), it does account for costs of labeling on non-labeled features.

As mentioned before, in the study of Overkott and Souza (2020), labeling the object when presented with a colored clip-art stimulus, for which the object served as the retrieval cue and color as the retrieved feature, produced a cost compared to a suppression condition. Mixture modeling showed that this cost was due to a decrease in the quantity of color information held in memory (both in categorical and continuous format) and some evidence also for a decrease in memory precision. This finding suggests that labeling may produce the involuntary filtering of the non-labeled object's features. This cost could be explained by the attentional-cue hypothesis: labeling the shape focused attention on this property at the expense of the color property.

The goal of the present study was to examine this possibility more closely. Before presenting the goals of the current study, we will review the literature on the storage of multi-feature objects in VWM and the selective encoding of some features over others.

10.2.3. Multi-Feature Objects in Visual Working Memory

VWM performance decreases as the number of visual features stored for a given item increases (Fougnie et al., 2010; Hardman & Cowan, 2015; Oberauer & Eichenberger, 2013; Olson & Jiang, 2002; Palmer et al., 2015; Quak et al., 2018; Swan et al., 2016). This does not mean that all features of an item are stored together. There is evidence that the visual features of an item are stored or retrieved independently from each other as reflected on uncorrelated errors between reports of multiple visual features of the same object (Bays et al., 2011; Fougnie & Alvarez, 2011; Schneegans & Bays, 2017; Shin & Ma, 2017).

Whether storage of all features of an item is obligatory or not has generated conflicting results. This research has investigated whether participants are able to only store one (relevant) feature of an item while filtering out the irrelevant features, thereby improving performance in comparison to conditions in which all features are relevant. One line of research has suggested that filtering of irrelevant information is constrained to whole items and not to their individual features (Bae & Flombaum, 2013; Foerster & Schneider, 2018; Gajewski & Brockmole, 2006; Gao et al., 2016; Luck & Vogel, 1997; Luria & Vogel, 2011; Marshall & Bays, 2013; Treisman & Zhang, 2006; Vogel et al., 2001). For example, Marshall and Bays, (2013) presented colored dots or oriented bars (single-feature items), or colored oriented bars (dual-feature items) as memoranda. The dual-feature items were used in two conditions: (1) participants were asked to retain both features in memory or (2) they were informed prior to the study phase which feature was relevant. The later condition allows participants to filter out the irrelevant feature, thereby promoting the retention of the relevant feature. However, performance in this filter condition was similar to the condition where both features were relevant and it was substantially worse than when only a single feature was retained in VWM – in line with the idea that feature information cannot be voluntarily filtered.

The other line of research has shown that item features can be stored independently at will depending on which feature is indicated as relevant to the memory test (Bocincova et al., 2017; Bocincova & Johnson, 2019; Chen & Wyble, 2015; Maniglia & Souza, 2020; Rock et al., 1992; Serences et al., 2009; Swan et al., 2016; Yu

& Shim, 2017). For example, Swan et al. (2016) asked participants to remember the color of an arrow (both color and orientation varied continuously), and report it using a color wheel in the memory test (see also, Shin & Ma, 2016), thereby making the color feature relevant and the orientation feature irrelevant. After half of the trials, a surprise memory test on the irrelevant orientation feature followed (see also, Rock et al., 1992). Thereafter participants were asked at random to either recall the orientation or color feature – meaning that both features were now relevant. In the surprise trial, orientation memory was poor and clearly worse than color memory on the previous trials. Mixture modeling revealed that this cost was particularly evident in the memory precision parameter, as precision of the orientation memory was low. Once participants were aware of both feature tests, orientation performance improved compared to when it was irrelevant. The later results suggest that selective attention to one of the features can gate encoding of only the attended information to VWM. In a pilot study in our lab (see details and data analysis on the OSF page: https://osf.io/z3yp3/), we also tested for the voluntary filtering of irrelevant features and whether labeling could assist in this process. Participants were presented with two sequential displays consisting of two items each (4 items in total). Following Marshall and Bays (2013), we presented items varying on one feature dimension (color or orientation; single-feature condition); or objects varying on both these features and this last presentation mode was split into a condition where both features were made relevant (dual-feature condition) or one feature was relevant and the other was irrelevant (filter-feature condition). Participants were instructed to label all relevant features or to perform suppression throughout. At test, memory of one of the relevant features was tested. Performance in the filter-feature condition was similar to the single-feature condition and a lot better than the dual-feature condition irrespective of labeling. These results support the view that participants are able to voluntarily filter the irrelevant feature information, in contradiction to the results of Marshall and Bays (2013).

In sum, recent evidence is mounting in support of the view that features can be voluntarily filtered. What is not known yet is whether this filtering process could occur involuntarily. To the best of our knowledge, no study has investigated whether features may be involuntarily filtered in VWM. This could happen, for example, when one labels only one feature of the object. Arguably labeling directs attention to this feature thereby boosting its encoding, consolidation, or maintenance in VWM. Could this come at the expense of the other relevant yet non-labeled features of the memory item? If labeling one feature leads to the involuntary filtering of the non-labeled features, this would indicate that the labeling boost involves suppression of non-labeled information to more effectively gate the entrance of the labeled feature in VWM and it would indicate that labeling induces trade-offs on how capacity is allocated in VWM.

10.2.4. The Present Study

Our main goal was to assess the fate of labeled and non-labeled features in VWM. In particular, we aimed to assess whether labeling of one feature would automatically lead to the filtering of non-labeled features. We investigated this question across three visual features: color, orientation, and spatial frequency.

The activation of categorical information hypothesis predicts that the labeling of an item feature increases continuous memory. This hypothesis is however silent regarding the impact of labeling on non-labeled features. One may wonder, however, whether activation of LTM may free VWM capacity, as it happens in the process of chunking (for example see, Chekaf et al., 2016; Nassar et al., 2018; Thalmann et al., 2019), in which case labeling one feature may actually free up capacity to encode the other feature in VWM leading to benefits for both labeled and non-labeled features. In contrast, the attentional-cue hypothesis predicts that labeling of one feature will increase attention to this feature at the expense of the processing of the other feature dimension, leading to costs to the non-labeled feature.

To test for these possibilities, we conducted two experiments. The general procedure across both experiments consisted of a delayed estimation task for memory items varying on two features (colored triangles; Experiment 1) or two out of three feature combinations (Gabor patches with color, orientation, and spatial frequency; Experiment 2). During item presentation, participants were asked to label one of the two item features, for example “green” for color, “left” for orientation, and “wide” for spatial frequency. Performance in these conditions was contrasted to a control condition, in which participants were asked to repeatedly say “bababa” aloud, thereby inhibiting verbal labeling. Finally, participants were then tested on both features of a randomly probed item. This testing procedure guaranteed that both features were relevant to the task, although only one of the features was labeled during study. This allowed us to assess the impact of labeling on the labeled and non-labeled features.

To foreshadow our results, we found asymmetric effects of labeling on the labeled and non-labeled features. We consistently observed a benefit for the labeled feature across all visual features studied (i.e., color, orientation, and spatial frequency) which was revealed by an increase in the storage of continuous information about these features, replicating and extending our previous research. However, the effects on the non-labeled feature were more varied: color was always involuntarily filtered when other visual dimensions were labeled, resulting in less continuous information about this feature in VWM. Orientation was not impacted by color labeling but suffered when spatial frequency was labeled. Lastly, spatial frequency showed no costs due to color or orientation labeling. These findings indicate that verbal labeling can be used to counteract the capacity limitations in VWM by increasing continuous information, but this benefit can come at the expense of the non-labeled feature.

10.3. Experiment 1

What happens when one feature of a multi-feature visual object is labeled? To address this question, in Experiments 1a and 1b, participants were presented colored triangles and they were requested to remember the color and orientation of all items. They completed this task under suppression, and under two labeling conditions that required them to either label only the colors or only the orientations of the items. At the memory test, both features of a single (randomly selected) item were reproduced. In Experiment 1a, a color wheel was presented for color reproduction. We then replaced the color wheel with a grey wheel (hidden color wheel) in Experiment 1b to rule out that the results in Experiment 1a could be explained by color wheel interference (Souza et al., 2016).

To foreshadow our results, verbal labeling improved recall performance for the labeled feature in contrast to suppression. Color memory improved when color was labeled but was impaired when orientation was labeled compared to the suppression condition. Orientation memory, in contrast, showed an improvement when orientation was labeled, and if at all, only small costs when color was labeled. These results indicate that some visual features are lost, or filtered, by labeling.

10.3.1. Methods

10.3.1.1. Participants

In total, 102 students from the University of Zurich were tested across Experiments 1a and 1b. Experiment 1a originally included a sample of 42 participants (*M* = 23.71; *SD* = 4.42; 25 women). Of these, five participants were excluded as they failed to follow the labeling instructions by either not labeling anything or giving the wrong type of label for the condition. Another person was excluded as the verbal label recording did not work. In total, the data of 36 participants were retained for the final analysis. Sample-size decision was as follows. We started with testing 30 participants. We then decided to increase the number of participants because the evidence obtained for the contrast of some of our conditions of interest (i.e., evidence for labeling effect on the labeled and non-labeled features) was ambiguous (i.e., Bayes Factor, BF, was between 0.33 and 3). To obtain clearer evidence for these contrasts, we added a second batch of 12 people (considering our counterbalancing across 6 participants).

Experiment 1b included a sample of 60 students (*M* = 22.87; *SD* = 3.91; 49 women). First, we obtained data of 36 participants to match the final dataset of Experiment 1a. As there was ambiguous evidence regarding a cost for the non-labeled feature in orientation recall, we increased the sample size to a total of 60 participants. One participant was excluded for not following the labeling instructions, leaving a total of 59 valid data-sets in Experiment 1b.

Only participants with German (or Swiss-German) mother tongue, aged between 18-35 years, and reporting normal color vision and normal or corrected-to-normal visual acuity could take part in the experiment. Participants had to sign an informed consent form prior to the study and were debriefed at the end. The experimental protocol was in accordance with the Institutional Review Board of the Psychology Institute from the University of Zurich and it did not require special approval.

10.3.1.2. Materials

All experiments were run in MATLAB using the Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997). Memory items were colored isosceles triangles. Colors were sampled from 360 values that varied on a continuous color circle defined in CIELAB color space, with L= 70, a= 20, b= 38, and radius= 60 (Zhang & Luck, 2008). The vertex of the isosceles triangle (30° angle) pointed in directions that varied in 360 degrees. Hereafter we will refer to it as the orientation feature. The length of the side of the triangle was set to 120 pixels. Memory items were presented within an imaginary circle with a radius of 200 pixels. The positions of the items were determined as follows: The position of the first item was randomly selected from 360°. The remaining three items were presented within a distance of 90°, 180° and 270° from the first item, thereby evenly spacing the memoranda, but varying their relative position from trial to trial. The items were presented in a sequence of two displays containing adjacent items. Thus, the two items presented on one screen were presented with a distance of 90 degrees. The memory items were presented against a grey background (RGB 128 128 128). In Experiment 1a, a color wheel was used for memory test, whereas in in Experiment 1b, a grey wheel (RGB 96 96 96) was presented.

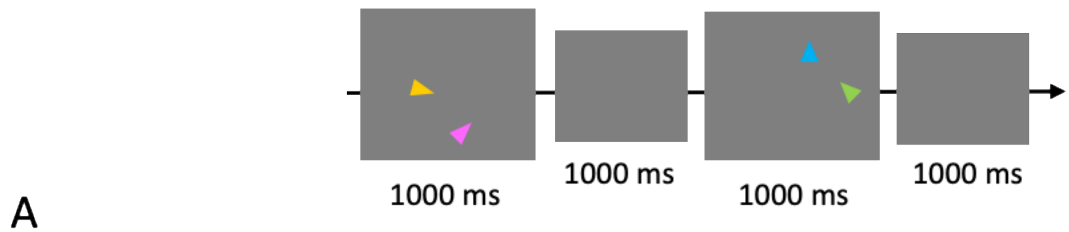
10.3.1.3. Procedure

At the beginning of every trial, a white (RGB 255 255 255) fixation cross was presented in the center of the screen for 1000 ms. Each display was then presented for 1000 ms, followed by a blank interval of 1000 ms (see Figure 1A). In total, four memory items were presented, two items for each display. During memory presentation, three labeling conditions were applied: participants were asked to either (a) label the color, (b) label the orientation, or (c) say “bababa” aloud (suppression condition). Their verbal output was recorded for offline check of compliance with the labeling instructions.

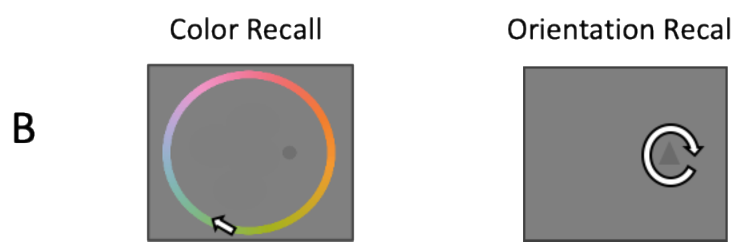
For the memory test (see Figure 1B), the orientation and the color of one item had to be recalled with the use of a continuous scale. The order in which the two features were tested was counterbalanced across trials. The test item (hereafter target) was randomly selected from the first or the second display for an equal number of trials, and a memory probe was shown at the target’s location. For the orientation test, a dark-grey (RGB 112 112 112) probe-triangle appeared at the location of the tested item with the mouse-cursor located in the probe’s center. The initial orientation of the probe-triangle was randomly chosen from the 360 possible angles. The probe rotated once participants started moving the mouse-cursor around, with the vertex of the triangle pointing in the same direction as the mouse cursor. For the color test, a dark-grey dot appeared at the location of the tested item together with a color wheel (in Experiment 1a) or a grey wheel (in Experiment 1b) and the mouse cursor. To change the color of the probe, participants moved the mouse cursor along the wheel which prompted the change in the probe’s color. Participants were instructed to adjust the orientation or the color of the probe as accurately as possible, and to click with the mouse to confirm their selection. After responding to both memory tests, a message indicating to press the spacebar to initiate the next trial appeared along with a reminder of the current labeling condition (e.g., say “ba ba ba” out loud now; label the colors; label the orientations).

Figure 1

Flow of the Experimental Procedure in Experiment 1 along with the Labeling Conditions



Suppression	“ba ba ba”	“ba ba ba”
Color	“orange, pink”	“blue, green”
Orientation	“right, up”	“up, left”



Note. Panel A shows the flow of one trial and Panel B shows the recall test display for color (Experiment 1a) and orientation recall, respectively.

Labeling conditions were completed in three separate blocks, and the order of these three blocks was counterbalanced across participants. In total, 4 practice trials and 80 test trials were completed in each block, resulting in 240 test trials.

10.3.1.4. Data Analysis

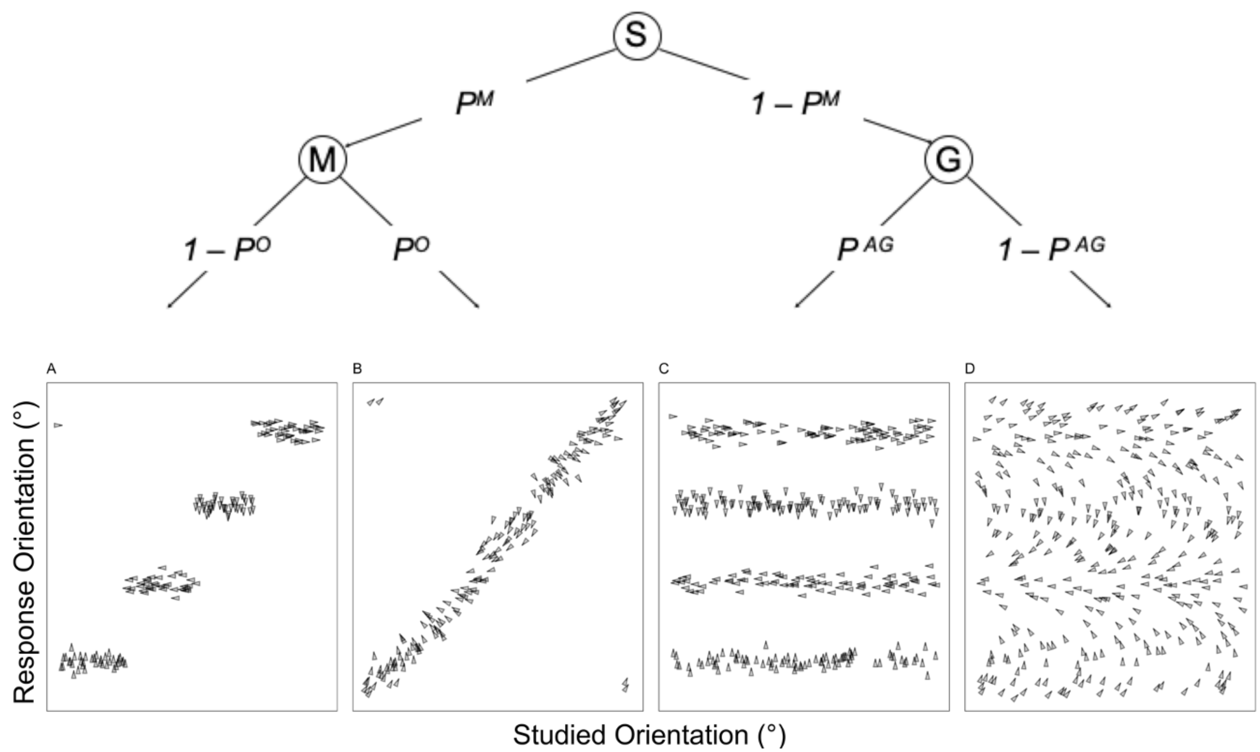
Recall Performance. Recall performance was assessed by calculating the deviation between the given response and the true value of the studied item in degrees.

The absolute value of the deviation can be taken as a model-free index of performance, referred to as recall error. We then submitted this data to a Bayesian ANOVA (hereafter BANOVA). The advantage of a Bayesian analysis is that it quantifies the strength of the evidence for both the null and the alternative hypothesis. One commonly employed measure is the Bayes Factor (BF). The BF is the strength of evidence for one hypothesis (e.g., the alternative) over another hypothesis (e.g., the Null), given the observed data. BFs should be interpreted as a continuous index of the strength of evidence in the data in support of one model over the other and provides the factor by which the ratio of our prior beliefs should be updated in light of the data. In other words, the BF can be reported in favor of the alternative (BF_{10}) or the null (BF_{01}), where $BF_{01} = (1/BF_{10})$. A BF_{10} larger than 1 gives evidence for the alternative hypothesis (i.e., for the presence of an effect), a BF_{10} lower than 1 provides evidence against an effect and hence evidence for the null hypothesis. A BF_{10} of 10 indicates that the alternative hypothesis is 10 times more likely than the null. Usually, BFs > 3 are regarded as providing substantial evidence for one hypothesis over the other, BFs > 10 as providing strong evidence, and BFs > 100 as providing decisive evidence (Jeffreys, 1961; Wetzels & Wagenmakers, 2012). We computed the BFs in line with Rouder et al. (2012) by using the BayesFactor package with default prior settings (Morey & Rouder, 2015) implemented in R (R Core Team, 2014).

Categorical-Continuous Mixture Modeling. We modeled participants' responses using the Bayesian hierarchical categorical-continuous mixture model of Hardman et al. (2017), which is illustrated in Figure 2 applied to the recall of orientations. The model assumes that responses about a stimulus (S) are either informed by memory (M; P^M) or reflect guessing (G; 1- P^M). Responses informed by memory can further be divided into categorical (1 - P^O) or continuous (P^O) information about the memory item. Panels 2A an 2B illustrate these two types of information by plotting the relation between the studied feature value against the recalled feature value. In Figure 2A, responses cluster around four canonical values (up, left, down, right) along the feature space representing only categorical knowledge. Figure 2B shows responses that vary continuously with the studied feature (thereby falling on a diagonal indicating covariation between the two). Continuous responses can be more or less fine-grained, thereby reflecting its continuous imprecision (σ^O). Very fine-grained responses are reflected by a dense diagonal line as depicted in Figure 2B, whereas less fine-grained responses lie around a broad diagonal line. When participants guess (G), they can do so by randomly selecting among the categories (P^{AG}), as illustrated in Figure 2C, or by randomly sampling one of the continuous values (1-P^{AG}), as shown in Figure 2D. In this mixture model, every category has a mean and standard deviation, which can be estimated freely by the model if no prior knowledge about the participants' categories is given. In the following experiments, the category means were freely estimated.

Figure 2

Categorical-Continuous Model Tree Exemplified for the Orientation Feature Space



Note. The upper part shows the model tree, where responses of a stimulus (S) can either be informed by memory (M; P^M) or reflect guessing (G; 1-P^M). Memory is divided into categorical responses (1-P^O) and continuous responses (P^O), which can be more or less fine-grained (σ^O). Guessing is divided into categorical (P^{AG}) or continuous guessing (1-P^{AG}). Panel A depicts categorical responses, here regarding four categories (up, right, down, left) are shown. Panel B shows continuous responses that align along a vertical line and a denser line reflects more precise responses. Panel C reflects categorical guessing, distributed along the orientation categories. Panel D shows uniform guessing.

For all experiments, we fitted the between-item model of the CatContModel package (Hardman, 2016) implemented in R. In this model variant, both categorical and continuous information relative to a memory item can be hold in memory at the same time. At test, however, the response is based on either categorical or continuous information. The within-item model variant, in contrast, assumes that both categorical and continuous information are integrated to inform response selection, but it has been reported to have worse model fit to the data of this task (Hardman et al., 2017; Souza & Skóra, 2017). Hierarchical models view the parameters of individual participants in a given condition as samples from a population-level distribution. The parameter values and distributional probabilities were determined through Markov chain Monte Carlo (MCMC) sampling techniques.

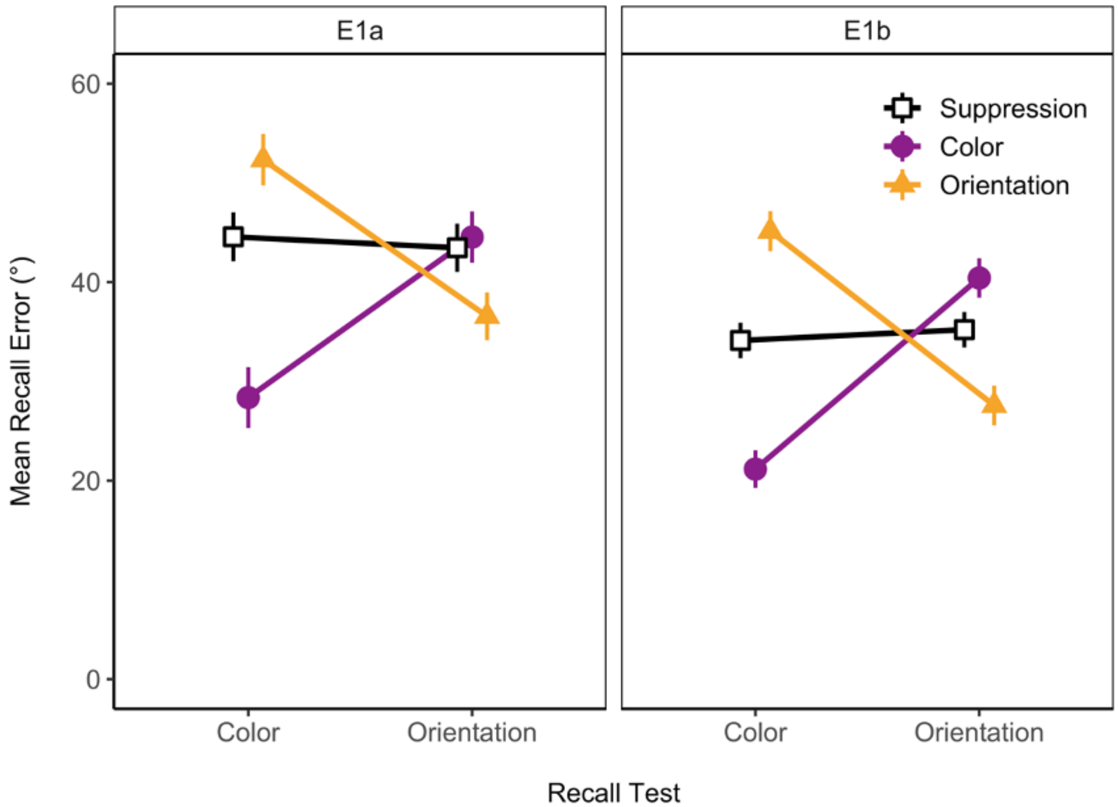
10.3.2. Results

10.3.2.1. Recall Performance

The aim of Experiment 1 was to investigate whether labeling one visual feature of an item boosts memory for this feature at the expense of the non-labeled features, indicating that these features were involuntarily filtered. Figure 3 shows the recall error across conditions. For both experiments, color recall improved (showing a smaller recall error) when color was labeled, and it showed an impairment when orientation was labeled, in contrast to suppression. Likewise, orientation recall improved when orientation was labeled compared to suppression; whereas when color was labeled, orientation memory remained either unaffected (Experiment 1a) or a small but credible cost was observed (Experiment 1b).

Figure 3

Mean recall error in degrees for Experiment 1a (E1a) and Experiment 1b (E1b)



Note. The error bars represent the 95% within-subjects confidence interval.

Experiment 1a. To estimate the labeling effect on recall error we first ran a two-way BANOVA with labeling condition (suppression vs. color label vs. orientation label) and recalled feature (color vs. orientation) as predictors. As shown in Table 1, the best model of the data included both main effects and their interaction, and there was overwhelming evidence to keep the interaction in the model. The interaction shows that labeling has different effects depending on whether the labeled or non-labeled feature is recalled.

Table 1

Bayes Factor (BF) of Models with Different Fixed Effects Over the Null and BF favoring the Best Model Over the Model Specified in Each Row (BF_{Best}/BF_{Mrow}) for Experiment 1a and Experiment 1b

		Included Fixed Effects				
Exp.	Model n°	Condition	Recall Test	Cond. x Test	BF_{10}	BF_{Best}/BF_{Mrow}
E1a	**1**	✓	✓	✓	$\mathbf{2.32 \times 10^{20}}$	**1**
	2	✓	✓	---	165.26	1.40×10^{18}
	3	✓	---	---	1.12×10^{3}	2.07×10^{17}
	4	---	✓	---	0.15	1.55×10^{21}
E1b	**1**	✓	✓	✓	$\mathbf{1.43 \times 10^{47}}$	**1**
	2	✓	✓	---	5.00	2.87×10^{46}
	3	✓	---	---	31.54	4.55×10^{45}
	4	---	✓	---	0.15	9.37×10^{47}

Note. ✓ = effect included in the model. Best model is printed in bold. Best model = model with higher BF over the Null.

To estimate the labeling benefits and costs, we further ran Bayesian *t*-tests separately contrasting each labeling condition to suppression. For color recall, there was evidence for a color labeling benefit ($BF_{10} = 9.27 \times 10^6$) and an orientation labeling cost ($BF_{10} = 815.19$) in contrast to suppression. For orientation recall, there was an orientation labeling benefit ($BF_{10} = 153.47$) and evidence against a color labeling effect ($BF_{10} = 0.21$).

Experiment 1b. In Experiment 1b, we replaced the color wheel by a grey wheel to assess whether the higher susceptibility of the color memory to labeling-induced filtering observed in Experiment 1a was due to color interference produced by the color wheel at the memory test. The color wheel shows all possible colors simultaneously, whereas in the orientation test, only a single orientation was shown at each time point Souza et al. (2016) demonstrated that the color wheel creates interference with memory retrieval, and that hiding the color wheel with a grey well, and then only reveling one color at a time at the probe location, improves color recall in VWM. By using the grey wheel, we equated the test interference between the color and orientation recall procedures. The same BANOVA applied to the data of Experiment 1b also revealed that the best model included both main effects and their interaction (see Table 1). Again, the presence of the interaction indicates that labeling one feature has different effects depending on whether the labeled or non-labeled feature is recalled.

We then estimated the labeling benefits and costs with Bayesian *t*-tests. For color recall, there was evidence for a color labeling benefit ($BF_{10} = 7.30 \times 10^{11}$) and for an orientation labeling cost ($BF_{10} = 2.86 \times 10^{8}$) in contrast to suppression. For orientation recall, there was evidence for an orientation labeling benefit ($BF_{10} = 4.77 \times 10^{4}$) and for a color labeling cost ($BF_{10} = 123.83$) in contrast to suppression. The finding of a color labeling cost for orientation memory stands in contrast to the result of Experiment 1a, where we found some evidence against a color labeling cost. The difference between the two experiments was that in Experiment 1b, the color wheel was hidden under a grey wheel to reduce color interference. Indeed, removal of the color wheel was associated with overall better performance particularly in the suppression condition (compare

Experiment 1a to Experiment 1b). Results of Experiment 1b therefore show that the higher susceptibility of color memory to involuntary filtering cannot be accounted by this feature being more prone to interference from the test situation. If anything, this modification made orientation memory more susceptible to labeling costs.

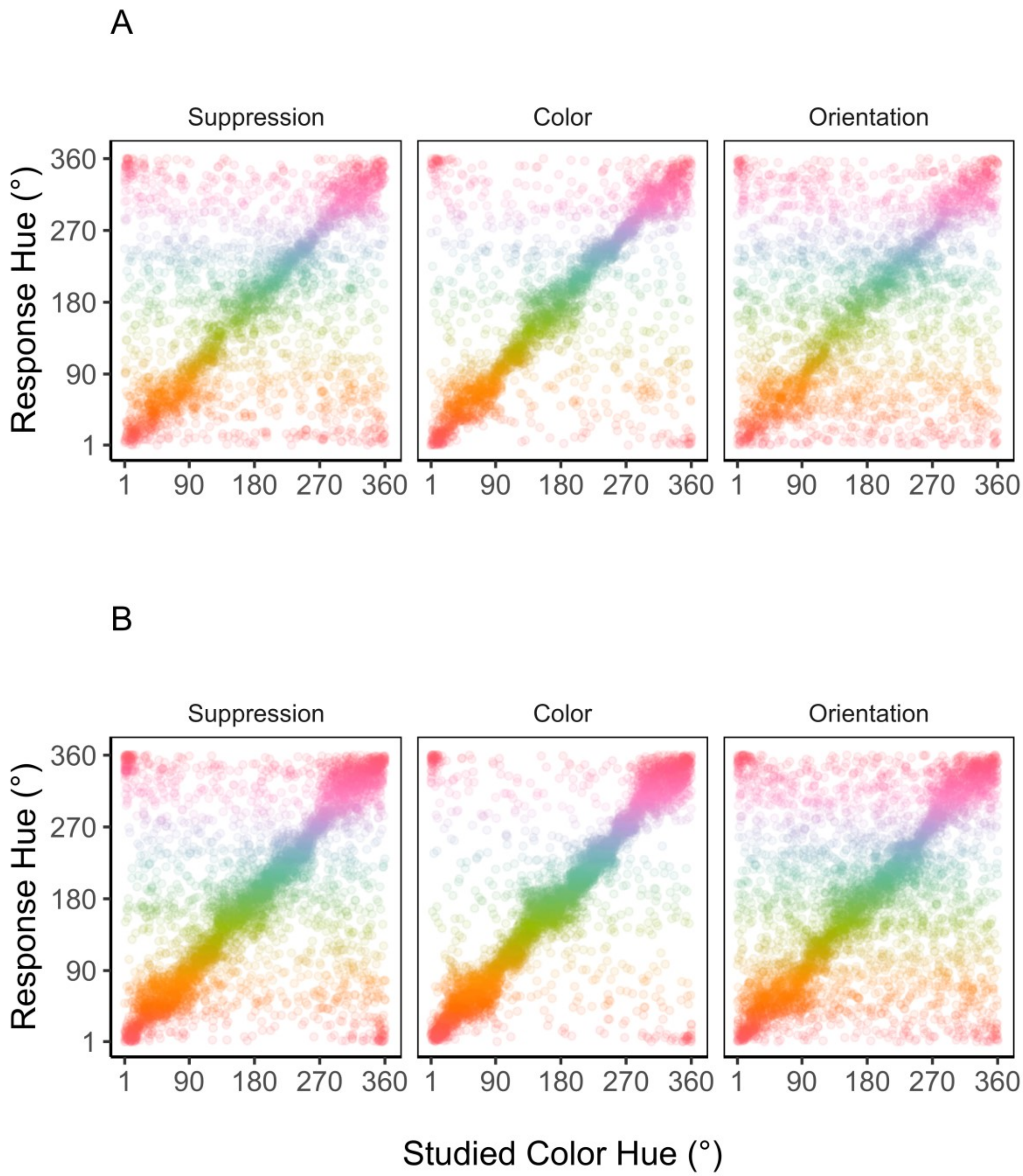
10.3.2.2. Categorical-Continuous Mixture Modeling

To get a first glance at the contribution of categorical and continuous responses to performance in the task, Figure 4 shows the scatterplot relating the studied feature to the response feature separately for the three labeling conditions and both types of recall tests. Figure 4A shows the color responses for Experiment 1a and Figure 4B shows the color responses in Experiment 1b. The scatterplots show a mixture of random guessing, continuous responses (diagonal), and also clusters of categorical responding that form some steps along the diagonal. For color recall, color labeling yielded a denser diagonal line in contrast to suppression and orientation labeling, indicating more continuous responses. Most guessing emerged for orientation labeling condition.

Figures 5A and 5B show scatterplots of orientation recall in Experiments 1a and 1b, respectively. Again, we can see random guessing and continuous responses along the diagonal. Categorical responding is less clear from this figure. Orientation recall is more densely distributed along the diagonal line for orientation labeling, indicating more continuous responses. Guessing behavior seemed to be similar for suppression and color labeling.

Figure 4

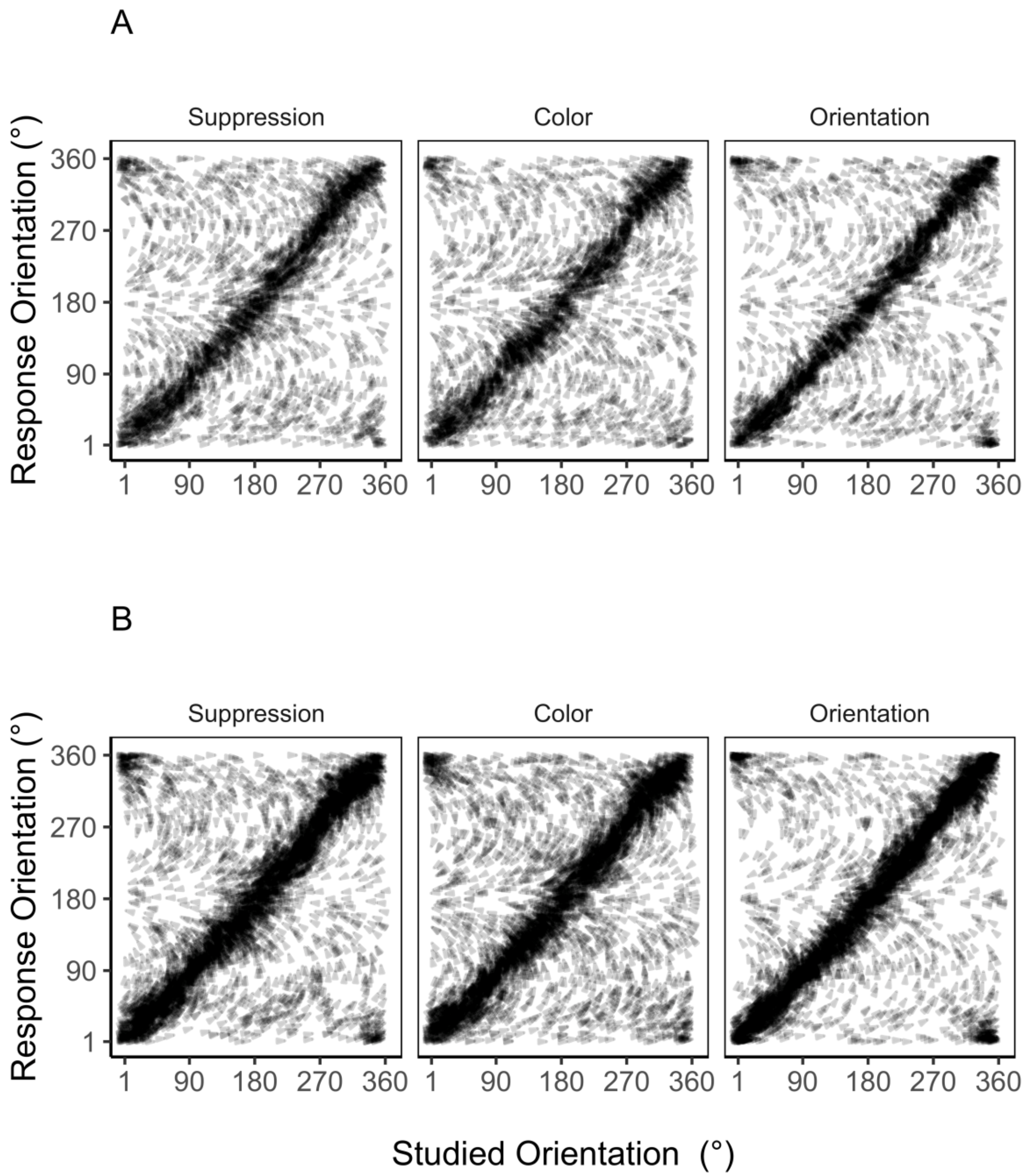
Scatterplots of Study-Response Distribution for Color in Experiment 1



Note. Panels A shows the response color hue as a function of studied color hue for the three labeling conditions in Experiment 1a and Panel B for Experiment 1b.

Figure 5

Scatterplots of Study-Response Distribution for Orientation in Experiment 1



Note. Panels A shows the response orientation as a function of the studied orientation for the three labeling conditions in Experiment 1a and Panel B for Experiment 1b.

We modeled the data of the three labeling conditions (i.e., suppression, color labeling, and orientation labeling) simultaneously in each experiment. However, we ran separate models for recall of color and orientation given that categorical biases in these two feature domains can differ. For each model, we ran 10,000 iterations, of which 2,000 were regarded as burn-in. First, we ran a full model allowing for an effect of labeling condition on all three parameters of the model. In subsequent steps, we removed the effect of labeling from each of the parameters and combinations of parameters as indicated in Table 2. For each model, we computed the Watanabe-Akaike Information Criterion (WAIC). The WAIC is used to assess the model fit based on the model predictive accuracy and includes a correction for the number of parameters used in the model (Gelman et al., 2014). This penalty term was helpful here as we constrained our models in regard to the number of parameters. The model with the smallest WAIC is the one that best explains the data.

As shown in Table 2, the best model for color recall in Experiment 1a included an effect of labeling on all three parameters of the model. For orientation recall, the best model did not include an effect of labeling condition on the continuous imprecision parameter. In Experiment 1b, for both color and orientation recall, the best model did not include an effect of labeling condition on continuous imprecision. However, for both recall tests, this model was only favored by a WAIC difference of 1 in comparison to the full model including all fixed predictors, which may indicate ambiguous evidence to exclude the effect of labeling on continuous imprecision.

Table 2

WAIC Comparison for all Fitted Mixture Models in Experiments 1a, 1b, and 2

		Color Model Labeling Effect on:					**Orientation Model** Labeling Effect on:				
		P^M	P^O	σ^O			P^M	P^O	σ^O		
Exp.	Model				WAIC	ΔWAIC				WAIC	ΔWAIC
E1a	1	✓	✓	✓	**89396**	**0**	✓	✓	✓	88838	11
	2		✓	✓	89747	351		✓	✓	88845	18
	3	✓		✓	89745	349	✓		✓	88882	55
	4	✓	✓		89400	4	✓	✓		**88827**	**0**
E1b	1	✓	✓	✓	140903	1	✓	✓	✓	141283	1
	2		✓	✓	141615	712		✓	✓	141442	160
	3	✓		✓	140919	16	✓		✓	141314	32
	4	✓	✓		**140902**	**0**	✓	✓		**141282**	**0**
E2	1	✓	✓	✓	**121250**	**0**	✓	✓	✓	**111647**	**0**
	2		✓	✓	121708	458		✓	✓	111724	74
	3	✓		✓	121250	0	✓		✓	111732	82
	4	✓	✓		121268	18	✓	✓		111650	3

Note. Δ indicates the difference score for this particular model in comparison to the best model. P^M = probability that information in in memory, P^O = probability of continuous information, σ^O = continuous imprecision.

We then assessed the group-level posterior estimates in each condition. We report the posterior estimates of the model including an effect of labeling on all parameters, even when this was not the best model because this allowed us to see the variability in the posterior of the parameters. Here, we report three parameters: categorical and continuous memory as well as continuous imprecision. To assess continuous memory, we calculated $P^M \times P^O$ – this value reflects the proportion of responses informed by continuous memory representations. Likewise, categorical memory was assessed by calculating $P^M \times (1\text{-}P^O)$ – reflecting the remaining proportion of memory responses that were informed by categorical information. To illustrate this, suppose that the model estimates that $P^M = 0.80$, and $P^O = 0.50$. This indicates that continuous memory representations informed 0.40 proportion of the responses, categorical memory informed 0.40 of the responses, whereas the remaining 0.20 of the responses reflected guessing. The continuous imprecision parameter (σ^O) was reported as outputted by the model, and it reflects the imprecision of the continuous memory. All reported models fitted the obtained data well (see Appendix A).

Table 3 presents group-level estimates for continuous memory, categorical memory, and memory imprecision in each labeling condition. Figure 6 presents posterior differences between each labeling condition in contrast to the suppression condition, thereby indicating how labeling modulated the retention of color and orientation information in mind. The posterior differences are presented with their mean (point) along with its highest density interval (HDI). The HDI reflects the range of values that covers 95% of the posterior. The zero represents no difference between the labeling condition and the suppression condition. When the HDI does not include zero, the labeling condition credibly differs from the suppression condition. Values above zero indicate a labeling benefit and values below 0 a labeling cost.

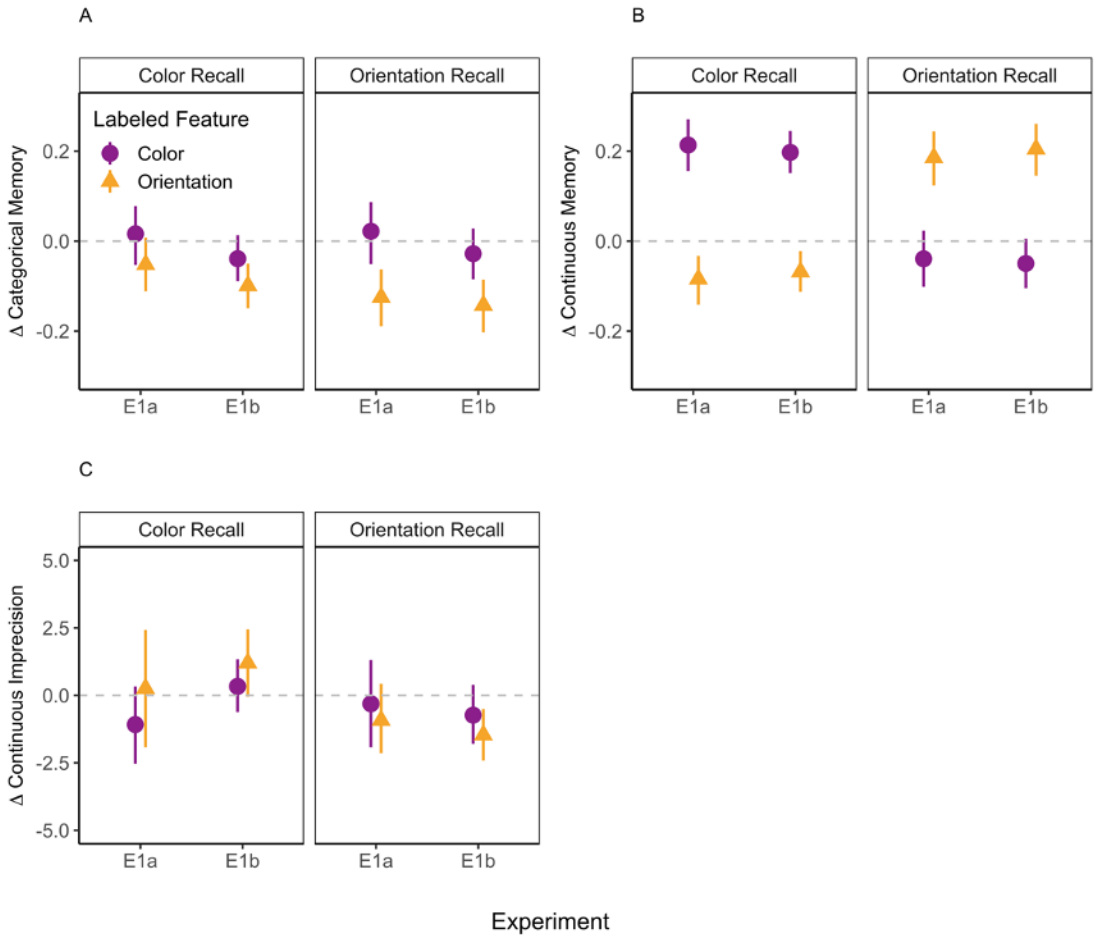
Figure 6A presents changes in categorical memory as a function of labeling. For color recall, color labeling did not change reliance on categorical representations, whereas orientation labeling tended to reduce categorical memory, but this effect was only credible in Experiment 1b. In contrast, for orientation recall, orientation labeling credibly reduced reliance in categorical memory in contrast to suppression, whereas color labeling had no credible effect.

Figure 6B presents changes in continuous memory as a function of labeling. For color recall, color labeling credibly increased continuous memory, whereas orientation labeling credibly reduced continuous memory in comparison to suppression. For orientation recall, orientation labeling credibly improved continuous memory compared to suppression, whereas color labeling tended to reduce continuous memory, but this effect was not fully credible in either experiment.

Figure 6C presents changes in continuous imprecision, reflecting the quality of continuous memory as a function of labeling. In line with the WAIC analysis, there were hardly any credible difference for continuous imprecision across the labeling conditions. Only for Experiment 1b, orientation memory showed a credible reduction in imprecision when orientation was labeled.

Figure 6

Changes of Categorical-Continuous Memory and Precision as a Function of Labeling for Experiment 1



Note. Dots depict the mean difference of the posterior distributions and the error bars depict the 95% HDI as a function of labeling. Panel A reflects the changes for the probability to retrieve categorical information, where the zero line presents no change of labeling color or orientation in contrast to suppression, Panel B for continuous information and Panel C the estimates of continuous memory imprecision.

Table 3

Posterior Means and Highest Density Intervals (HDI) of the Full Models in Experiments 1a, 1b, and 2

		Memory for Color						Memory for Orientation					
		Categorical		Continuous		Cont. Imprecision		Categorical		Continuous		Cont. Imprecision	
Exp. + Condition		Mean	95 % HDI	Mean	95 % HDI	Mean	95 % HDI	Mean	95 % HDI	Mean	95 % HDI	Mean	95 % HDI
E1a													
	Suppression	0.32	[0.25-0.39]	0.29	[0.22-0.35]	11.78	[10.46-13.23]	0.24	[0.17-0.32]	0.41	[0.33-0.49]	13.58	[12.26-14.88]
	Color	0.34	[0.27-0.41]	0.50	[0.43-0.57]	10.70	[9.74-11.71]	0.27	[0.19-0.35]	0.37	[0.29-0.45]	13.27	[11.87-14.71]
	Orientation	0.27	[0.20-0.35]	0.20	[0.14-0.27]	12.04	[9.85-14.22]	0.12	[0.06-0.17]	0.60	[0.52-0.67]	12.66	[11.66-13.67]
E1b													
	Suppression	0.40	[0.35-0.45]	0.37	[0.32-0.42]	11.88	[11.17-12.62]	0.29	[0.22-0.35]	0.48	[0.41-0.54]	15.24	[14.25-16.35]
	Color	0.37	[0.31-0.42]	0.56	[0.50-0.62]	11.88	[11.17-12.62]	0.26	[0.20-0.32]	0.43	[0.36-0.49]	14.51	[13.37-15.64]
	Orientation	0.33	[0.27-0.37]	0.28	[0.23-0.32]	11.88	[11.17-12.62]	0.14	[0.09-0.20]	0.68	[0.62-0.74]	13.77	[12.88-14.74]
E2	Suppression	0.48	[0.40-0.56]	0.32	[0.25-0.39]	15.61	[13.88-17.33]	0.43	[0.36-0.50]	0.19	[0.15-0.24]	7.86	[6.94-8.78]
	Frequency	0.43	[0.33-0.52]	0.26	[0.19-0.33]	15.87	[13.85-17.82]	0.41	[0.33-0.48]	0.14	[0.10-0.19]	9.62	[8.29-10.89]
	Color	0.55	[0.48-0.62]	0.44	[0.37-0.51]	12.91	[11.69-14.26]	---	---	---	---	---	---
	Orientation	---	---	---	---	---	---	0.26	[0.20-0.32]	0.44	[0.37-0.50]	7.85	[7.17-8.52]

10.3.3. Discussion

In Experiment 1 we showed that labeling one feature of a multi-feature object can have two types of consequences: (1) it increases the amount of continuous information one can retain in mind about the labeled feature as reflected in an increased probability of continuous memory recall, and (2) this boost may come at the expense of the retention of categorical and continuous information about the non-labeled feature. This was the case for color memory, but not credibly for orientation memory. More specifically, we found that labeling orientation led to some cost for categorical color memory (see Exp. 1b) and a very credible cost in continuous color memory (Exps. 1a and 1b) - indicating that fine-grained information about the color feature was involuntarily filtered when orientation was labeled. In contrast, color labeling did not lead to a credible cost for either categorical or continuous information about orientation as revealed by the parameters in the mixture model. There was, however, a credible, albeit small, cost for color labeling on orientation recall error in Experiment 1b. These results suggest that orientation memory was less likely to be involuntarily filtered than color memory. We also showed that this higher cost for color memory was not explained by color memory becoming more susceptible to color interference at test, as reducing color interference at test by using a grey wheel in Experiment 1b did not change the pattern of results.

These asymmetric effects of color and orientation labeling on recall performance indicate that some visual features may be involuntarily filtered when other features are labeled. So far, however, we do not know whether involuntary filtering of visual features is the norm and orientation is an exception, or whether the reverse is true with color information being particularly vulnerable to filtering due to labeling.

10.4. Experiment 2

The goal of Experiment 2 was to test the impact of labeling other visual features besides color and orientation. This experiment used Gabor patches that could vary in three visual dimensions (i.e., spatial frequency, color, and orientation). By adding the third visual feature (namely spatial frequency), we aimed to test whether labeling this feature could boost its retention in memory at the expense of other visual features, and whether labeling the other features (color or orientation) would lead to a cost for spatial frequency memory. We aimed to assess the likelihood of two hypotheses: (H1) = Labeling enhances memory for the labeled feature at the expense of the non-labeled feature; (H2) = labeling enhances memory for the labeled feature with no costs for the non-labeled features. Or, a mix of the two depending on the visual feature. We also hoped this would provide further insight regarding to what types of visual features are susceptible to a labeling cost.

We preregistered our hypotheses (https://osf.io/2spwt/) [12]: we expected to replicate the results obtained in Experiment 1, namely that labeling spatial frequency improves memory for this feature, it impairs memory for color, and it has no impact on memory for orientation. Critically, by assessing the impact of labeling color or orientation on memory for spatial frequency, this would allow us to assess whether H1 is true (and orientation is a special case) or whether H2 is more likely to be true (and color is a special case).

[12] In the preregistration, we mentioned that the results of Experiment 1a and 1b were similar. This was based on $N = 36$ participants tested in Experiment 1b, but we decided to later test up to 60 participants to determine whether the cost for orientation memory in recall error was credible.

10.4.1. Methods

10.4.1.1. Participants

In total a new sample of 61 students of the University of Zurich were tested in Experiment 2 (*M* = 23.97; *SD* = 4.18; 45 women). Participants completed two 1-hour sessions. In the preregistration, we mentioned to start data collection with a sample of 24 participants, and that we would add more participants until we reached BFs ≥ 10 for comparison of our conditions of interest or that we would stop data collection once we have collected a total of 60 participants. The latter was our key determinant to stop data collection.

Seven participants were excluded due to not following the labeling instructions [13], one for not attending the second session, one for aborting the experiment in the second session, and one because they admitted to the experimenter after the experiment was over that they did not understood the instructions for labeling and just repeated the terms appearing in the instructions. Thus, the final data set submitted to the analysis consisted of 51 participants.

Participants fulfilled the same inclusion criteria as in Experiment 1, except for two participants who did not inform us prior to the study that their mother tongue was not German. As their German was sufficiently good to label the features, we included their data into the analysis. Participants were exposed to the same experimental protocol as in the previous experiments.

[13] Five of these participants additionally labeled the other feature during the frequency labeling condition, and the other two participants labeled the correct feature on less than 70% of occasions.

10.4.1.2. Materials and Procedure

In Experiment 2 (see Figure 7A), three Gabor patches were presented sequentially against a black background (RGB 0 0 0), with each Gabor remaining onscreen for 1000 ms, followed by an inter-item interval of 1000 ms. The envelope of each Gabor patch was defined with a size of 61 pixels and a radius of 61 pixels. The Gabor sigma was set to 10.17. The background of the Gabor was black (RGBA 0 0 0 0) with a pre-contrast multiplier of 1.0. The three Gabor patches were presented equally spaced in an imaginary circle centered in the middle of the screen. The exact locations of the items varied from trial to trial. The Gabor spatial frequencies ranged from 12 pixels/cycle to 24 pixels/cycle (0.19° and 0.39°) in 13 steps.

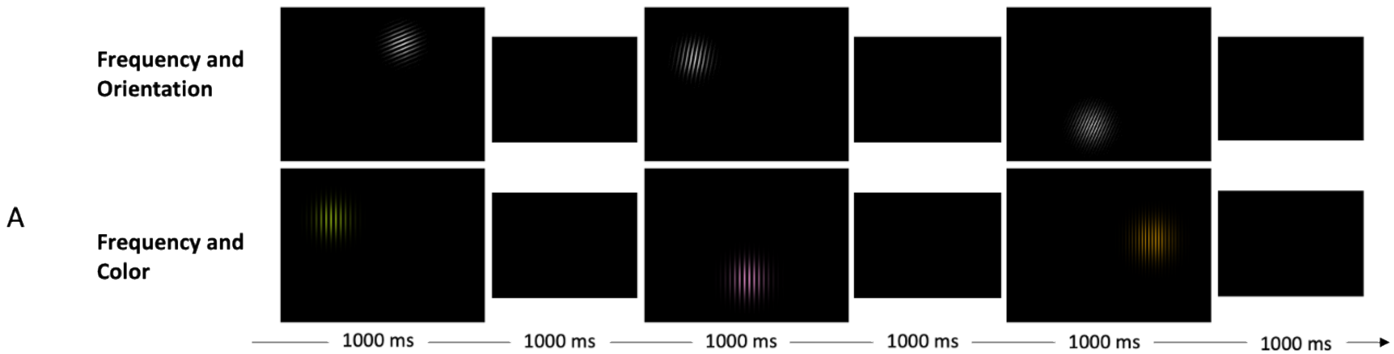
Participants completed two sessions. In one session, they were presented with Gabor patches that varied in spatial frequency (1 out of 13 values) and orientation (0-180°), whereas color was fixed at a single value (e.g., white). In the other session, the Gabor patches varied in spatial frequency and color (0-360 colors as defined in Experiment 1), whereas orientation was fixed (0°). Hence although the items contained three features, only two features were varied at a time.

As for the critical labeling manipulation, participants were asked to perform the task under suppression, or they were required to label aloud the spatial frequency, or the other feature of the item (either color or orientation). We did not assess the impact of orientation labeling on color memory and vice versa because we had addressed this issue in Experiments 1a and 1b, and we aimed to maximize the number of trials on the new conditions that addressed the new experimental questions posed in Experiment 2. The labeling conditions were implemented in three blocks whose order was counterbalanced across participants. Each block consisted of 78 trials, resulting in 234 trials in total for one session. Before the start of each block, participants completed three practice trials. Participants started each trial by pressing the space bar. Before pressing the space bar, they were remembered of the current labeling condition (e.g., say “ba ba ba” out loud now; label the spatial frequencies, label the colors, or label the orientations). The verbal output during the study phase was recorded for offline check of compliance with the labeling instructions. In this experiment, we coded each of the verbal responses to assess which labels were applied to which memoranda.

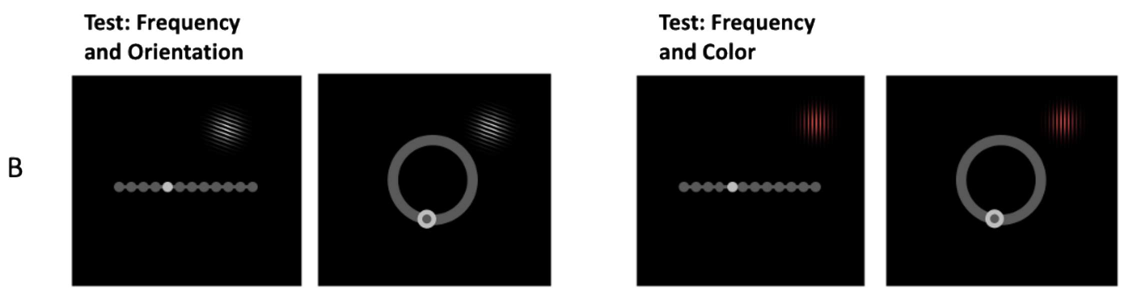
At test (see Figure 7B), one item was randomly chosen as the test target and a memory probe was shown at the target’s location. The probe was shown with a randomly selected spatial frequency. For the session in which orientation was varied, a random orientation was selected for the probe, and color was always white. For the session in which color was varied, a random color was selected for the probe, and orientation was fixed at 0°. Participants were requested to reconstruct both relevant features of the target item. The order in which the two features were tested was counterbalanced across trials. Participants adjusted the spatial frequency of the probe by moving a dot (RGB 150 150 150) on a dark grey slider (RGB 96 96 96) presented in the middle of the screen. They adjusted the orientation or color of the probe by moving a dot on a grey wheel presented in the middle of the screen. When participants were satisfied with the adjusted featured, they confirmed their response by a left-mouse klick.

Figure 7

Example of One Trial in Experiment 2 for Frequency-Orientation and Frequency-Color Session



Articulatory Suppression	«ba ba ba»	«ba ba ba»	«ba ba ba»
Frequency Labeling	«wide»	«middle»	«narrow»
Orientation Labeling	«diagonal/horizontal»	«vertical»	«diagnoal»
Color Labeling	«green»	«violet»	«orange»



Note. Panel A presents the presentation phase of the Gabor patches for the two session conditions (frequency-orientation and frequency-color). Panel B gives an example of the memory test for these two sessions.

10.4.2. Results

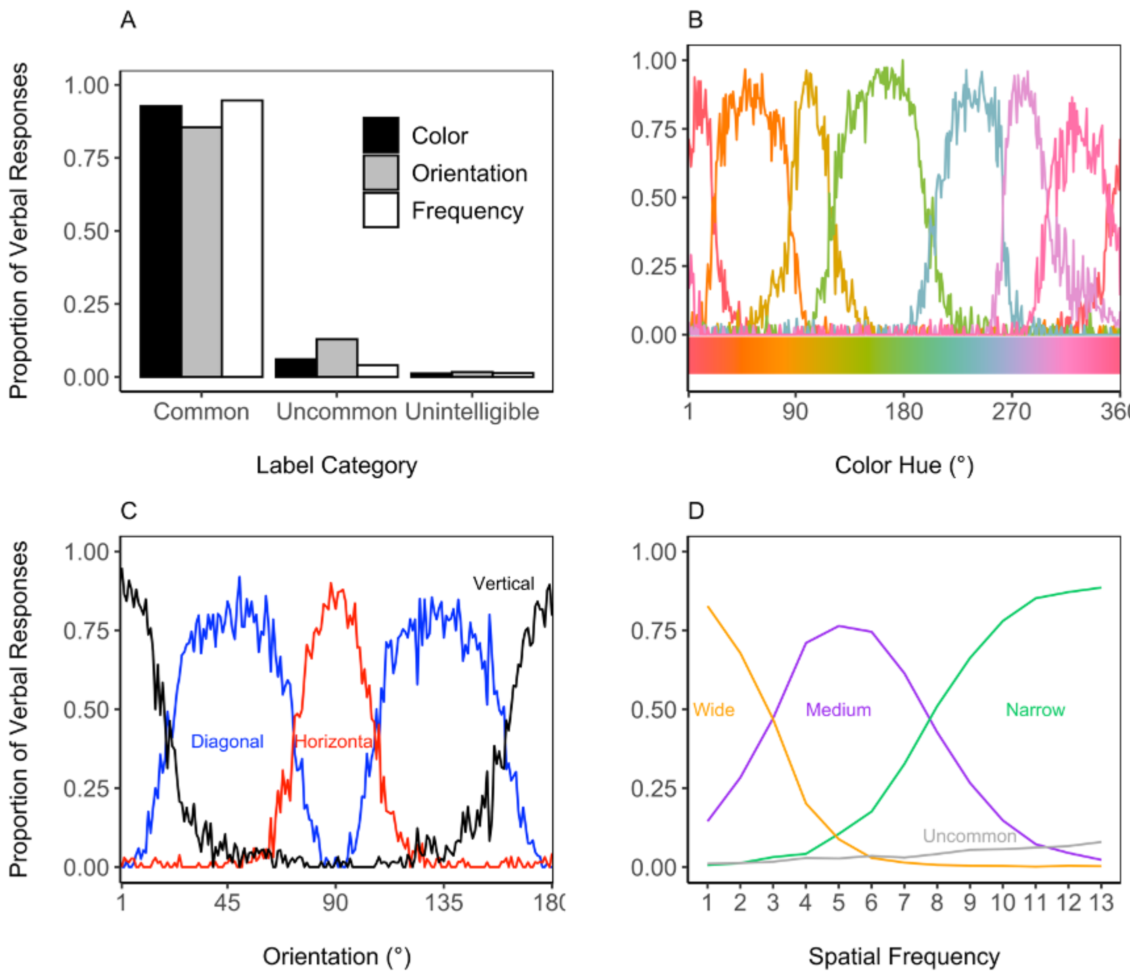
10.4.2.1. Verbal Labeling Output

We recorded the verbal responses during the study phase of the working memory task in all our experiments. However, in Experiment 1, two memory items were simultaneously presented, which did not allow for a direct inference of which label was assigned to which memory item. In Experiment 2, memory items were sequentially presented, which allowed us to further analyze the verbal labeling output data to assess the variety of labels applied to the colors, orientations, and spatial frequency of each Gabor patch, and which feature values the labels were applied to.

Participants used a total of 90 different color labels, 93 orientation labels, and 31 frequency labels. The majority of the color labels belonged to a set of seven basic color categories (e.g., red, orange, yellow, green, blue, purple, and pink) - hereafter referred as common category, as opposed to the usage of more uncommon labels (e.g., turquoise, yellow-green, dark orange, blueish), or unintelligible responses. Likewise, three terms were commonly used for orientation (e.g., diagonal, horizontal, vertical), and three terms were used for spatial frequency (e.g., wide, thin, medium). The proportion of occasions in which these sets of 7 color terms, 3 orientation terms, and 3 spatial frequency terms were used (hereafter common category) is depicted in Figure 8A. Other terms that did not belong to this set were classified as uncommon, and we also coded for unintelligible responses (output was not understandable or the participants remained silent). Based on this classification, it is clear that more labels were assigned to the color space than to the orientation and frequency space. However, overall participants used the common category on the majority of trials, and this was similar across the features.

Figure 8

Analysis of the Labels used by the Participants in Experiment 2



Note. Panel A shows the proportion of occasions labels within the categories of common, uncommon, and unintelligible were applied for color, spatial frequency, and orientation. Panel B shows the proportion of times one of the seven common color labels was used to refer to a given color on the wheel (as shown in the x-axis). A proportion of 1 indicates that the x color on the wheel was labeled with the same label by all participants. The lower the proportion, the less often participants used that label to refer to that given color. Each color term is represented by the line with its prototypical color. Similarly, Panel C shows the proportion of times one of the three common orientation labels was used by the participants to refer to the different orientations. Panel D shows the proportion of times each of the three common spatial frequency labels (and the uncommon labels) were applied to each of the spatial frequency values used in the study.

Figure 8B presents the proportion of occasions one of the seven basic color labels was used (across all participants) to refer to the 360 colors. There was high agreement between participants regarding the labeling of the colors during the VWM trials. The same approach was used for the orientation labels, which ranged from 1 to 180 values. Figure 8C shows that for orientation, three broad labels were used across the orientation space. Figure 8D shows that three broad labels were used for frequency labeling. We further plotted the proportion of the uncommon labels, which is distributed in close proximity to the x-axis, showing that these labels were not systematically applied to a section of the spatial frequency continuum.

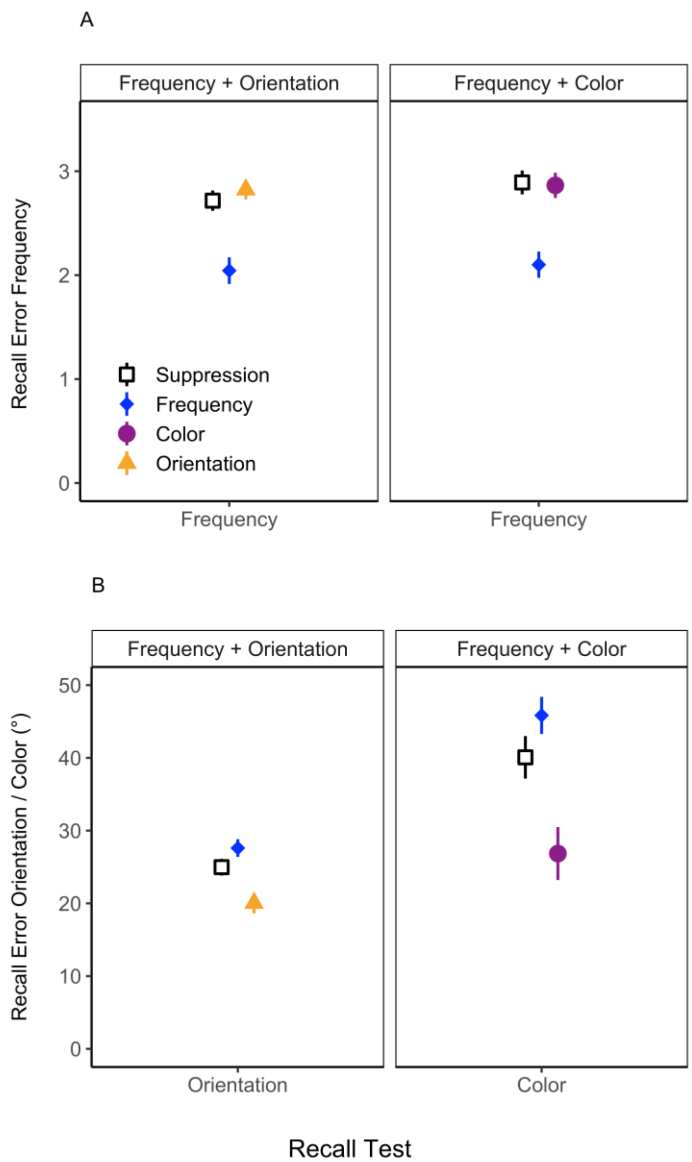
10.4.2.2. Recall Performance

For all features, we computed a measure of recall error by computing the absolute difference between the true feature value of the item and the participant's response. For spatial frequency, this measure ranged between 1 and 12 pixels/cycle. For orientation, this measure ranged between 0° and 90°, and for color between 0° and 180°. We evaluated the effect of the labeling condition (e.g., suppression, labeling the reported feature, labeling the other feature) upon each of these measures separately.

Figure 9A shows the error in recalling spatial frequency and Figure 9B the error for recalling orientation and color. Figure 9A shows that labeling spatial frequency reduced the error in recalling this feature compared to suppression, whereas labeling the color or the orientation of the Gabor had no credible impact. This shows that labeling boosted spatial frequency memory, and that information about spatial frequency was not lost when participants labeled the other features.

Figure 9

Mean Recall Error in Experiment 2 as a Function of Labeling Condition and Recalled Feature



Note. Panel A shows the mean recall error for frequency recall. The error is presented for the two sessions, frequency + color and frequency + orientation recall test along with the labeling conditions. Panel B shows mean recall error for orientation and color as a function of labeling condition, for their respective sessions. Note that recall error varied from 1-90° for orientation recall and from 1-180° for color recall. Error bars represent the 95% within-subjects confidence interval.

For orientation memory (Figure 9B), orientation labeling reduced recall error, whereas spatial frequency labeling increased recall error compared to suppression. Likewise, for color memory, color labeling reduced recall error, whereas spatial frequency labeling increased recall error compared to suppression. These results indicate that orientation and color information were lost or involuntarily filtered when spatial frequency was labeled.

We contrasted recall error in each labeling condition to the one observed in the respective suppression condition using Bayesian *t*-tests (see Table 4 [14]). Verbal labeling benefitted all three labeling conditions in contrast to suppression. For spatial frequency memory, there was inconclusive evidence whether labeling orientation led to a cost, and evidence for no cost when color was labeled. In contrast, spatial frequency labeling led to a cost for the non-labeled feature regardless of whether the non-labeled feature was color or orientation.

Table 4

BFs in favor (BF_{10}) of a Labeling Benefit or Cost in Experiment 2

	Frequency + Orientation Test Recall:		Frequency + Color Test Recall:	
	Frequency	Orientation	Frequency	Color
Labeling				
Frequency	5.47×10^6	20.40	5.19×10^9	25.78
Orientation	0.70	8.62×10^3	---	---
Color	---	---	0.16	7.68×10^3

Note. Green font indicates a labeling benefit or evidence against a cost, whereas red font indicates a labeling cost. Black font indicates ambiguous evidence.

[14] The Bayesian *t*-tests were not preregistered.

In the preregistration we mentioned to submit the data to a BANOVA. For this purpose, we calculated *z*-score values to directly compare the three different types of feature recall using the same scale. The *z*-scores were computed for each of the four recall tests depicted in Figure 9 by subtracting the mean recall error for that type of test averaged across all labeling conditions, divided by the standard deviation.

We first ran a 3-way BANOVA on the *z*-scored recall error with labeling (color, orientation, frequency), test condition (frequency-orientation, frequency-color) and tested feature (orientation, color, frequency) as fixed predictors, and subject as random predictor. The best model included labeling condition and tested feature as well as their interaction into the model (BF_{10}= 1.86×10^{47}). This model was favored against the second-best model including condition, test condition, tested feature and the interaction of condition × tested feature by a $BF_{10} = 6.04$.

To estimate more closely the effect of verbal labeling on the recall test for the two test conditions we ran two independent BANOVAs for the frequency-orientation and frequency-color conditions recall having labeling and tested feature as predictors (Table 5). For both BANOVAs, the best model included both main effects and their interaction and the inclusion of the interaction was clearly favored.

Table 5

Bayes Factor (BF) of Models with Different Fixed Effects Over the Null and BF favoring the Best Model Over the Model Specified in Each Row (BF_{Best}/BF_{Mrow}) for Experiment 2

		Included Fixed Effects				
Test Condition	Model n°	Condition	Recall Test	Cond. x Test	BF_{10}	BF_{Best}/BF_{Mrow}
Frequency-Orientation	**1**	✓	✓	✓	**2.37×10^{23}**	**1**
	2	✓	✓	---	0.58	4.11×10^{23}
	3	✓	---	---	4.54	5.21×10^{22}
	4	---	✓	---	0.13	1.88×10^{24}
Frequency-Color-	**1**	✓	✓	✓	**4.34×10^{24}**	**1**
	2	✓	✓	---	9.71	4.46×10^{23}
	3	✓	---	---	78.63	5.51×10^{22}
	4	---	✓	---	0.13	3.46×10^{25}

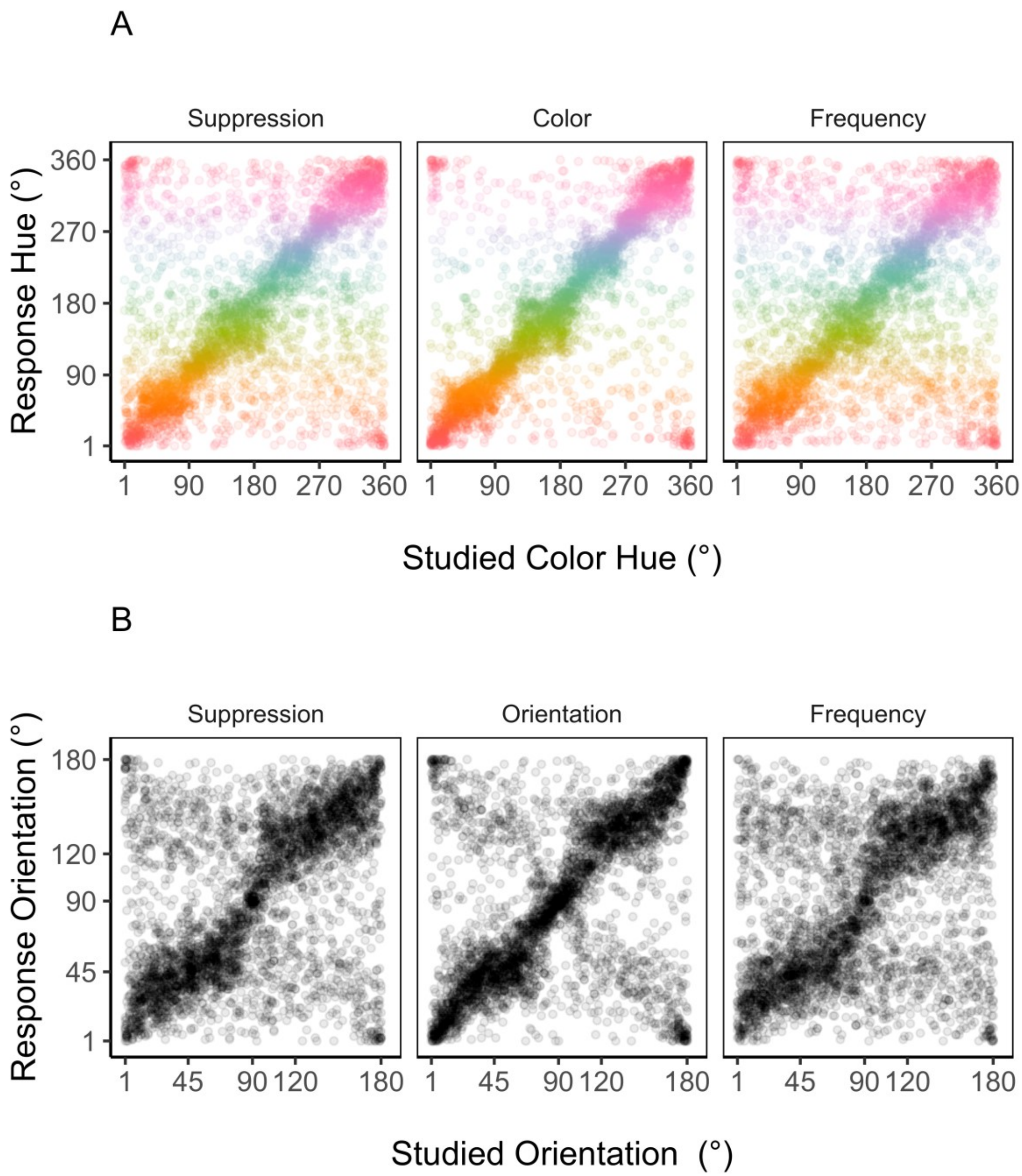
Note. ✓ = effect included in the model. Best model is printed in bold. Best model = model with higher BF over the Null.

10.4.2.3. Categorical-Continuous Mixture Modeling

Figure 10 shows the scatterplot of participants' responses against the studied color hue or orientation for all labeling conditions in Experiment 2. Similarly to Experiment 1, we can observe random guessing, continuous memory responses (diagonal), and categorical clusters along the diagonal for both color and orientation.

Figure 10

Scatterplot of Studied Feature Against Participants' Responses



Note. Panel A shows the studied color hue plotted against participants' responses for the three labeling conditions and Panel B for the orientation feature.

To estimate how verbal labeling affected continuous and categorical information held in memory for the continuous feature dimensions of color and orientation, we submitted participants' responses to the CatCont mixture model, as done in the previous experiment [15]. Note that again, we separately modeled recall of color and orientation given that categorical biases are different in these feature dimensions, but modeled all labeling conditions simultaneously. For each model, we ran 10,000 iterations, of which 2,000 were regarded as burn-in. We first ran the full model, containing the fixed effect of labeling condition on all parameters. We then constrained this full model as done in Experiment 1. Table 2 presents all models, alongside their WAICs, and their relative comparison. For color recall, two models yielded the same WAIC. One of them was the full model, and we decided in favor of this model to be more conservative. For orientation recall, the full model was the best model. Note that the model for orientation recall fitted the obtained data less well (see Appendix A).

As done in Experiment 1, based on P^M and P^O we calculated the probability of retrieving categorical and continuous memory, and these estimates are presented in Table 3 along with the group-level estimates for continuous imprecision. We then computed posterior differences between the labeling conditions compared to the suppression condition, which are displayed in Figure 11. To recapitulate, values above zero indicate a labeling benefit and values below 0 a labeling cost.

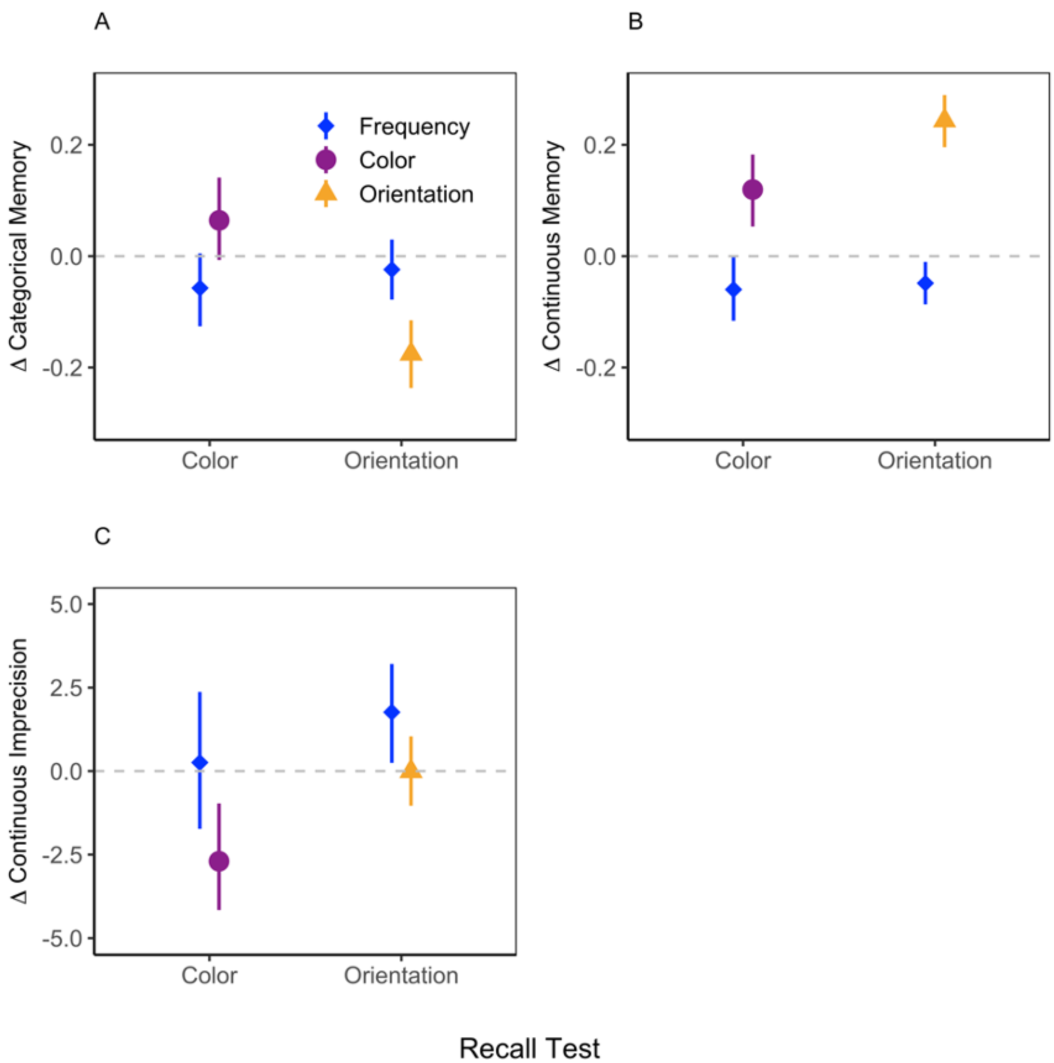
Figure 11A shows the changes in categorical memory as a function of labeling. For color recall, categorical memory tended to increase when color was labeled, and

[15] The mixture modeling part was not preregistered.

decrease when spatial frequency was labeled, but these effects were not fully credible. In contrast, for orientation recall, orientation labeling decreased categorical memory compared to suppression, but spatial frequency labeling had no credible effect. Figure 11B shows that color labeling increased continuous color memory in contrast to suppression. Spatial frequency labeling tended to decrease continuous color memory, although not credibly as 0 is within the HDI. For orientation recall, continuous memory clearly increased under orientation labeling, whereas spatial frequency labeling tended to decrease continuous orientation memory. Figure 11C shows that continuous color imprecision decreased when color was labeled compared to suppression, whereas spatial frequency labeling had no credible impact. For continuous orientation imprecision, orientation labeling had no effect whereas spatial frequency labeling produced a credible increase in imprecision, showing a cost.

Figure 11

Changes on Categorical, Continuous Memory, and Continuous Imprecision as a Function of Labeling for Experiment 2



Note. As in Experiment 1, dots depict the mean difference of the posterior distributions as a function of labeling and the error bars depict the 95% HDI. The dotted line represent no change in relation to the suppression condition.

10.4.3. Discussion

In Experiment 2 we assessed the impact of labeling other visual features of the memory items. We used Gabor patches that could vary in three dimensions (namely, spatial frequency, color, and orientation). We aimed to assess the likelihood that labeling enhanced memory for the labeled feature at the expense of the non-labeled feature (H1) or whether there was no cost for the non-labeled feature (H2). We found evidence for both hypotheses, suggesting that visual features are differently affected by the verbal labeling of other features. Color labeling did not yield a cost for frequency recall, suggesting that the frequency feature was not filtered. This finding extended the results of Experiment 1, where color labeling did not yield a consistent cost for orientation recall. In Experiment 2, there was ambiguous evidence to whether orientation labeling produced a cost for recalling spatial frequency. This is in contrast to Experiment 1, where we found that orientation labeling led to the filtering of continuous color information. Labeling the third feature, conversely, namely spatial frequency, led to a cost in recalling both color and orientation.

To summarize, in Experiment 2 we again found consistent benefits for recall of the labeled features, but more mixed patterns of costs for recall of the non-labeled features. Labeling the spatial frequency led to some loss of color and orientation information (labeling cost). This suggests that both the color and orientation feature can be involuntarily filtered, thereby extending the findings of Experiment 1 with a new visual feature. Moreover, we found across both experiments that the color feature was always filtered when other features were labeled, whereas color labeling did not lead to the filtering of the other features. These results show that color is one of the most vulnerable features to involuntary filtering.

10.5. General Discussion

Verbal labeling of continuously-varying colors and shapes in delayed estimation tasks have been found to produce benefits for VWM (Overkott & Souza, 2020; Souza et al., 2020; Souza & Skóra, 2017). Here we extended this finding to two additional visual features, namely orientation and spatial frequency. Critically, mixture modeling indicated that this labeling benefit originated from an increase in continuous memory, meaning that verbal labeling increased the retention of fine-grained information about the labeled visual features. This stands in contrast to prior assumptions that labeling would only provide categorical information about an item (Donkin et al., 2015; Hardman et al., 2017). Here, we addressed the question of what happens to the non-labeled features of an item. In one previous study (Overkott & Souza, 2020), we observed that labeling the shape of objects produced a cost to the retention of color information. This begged the question of whether the labeling boost would come at a cost to the retention of other features of the memory item.

10.5.1. Costs of Verbal Labeling to Non-Labeled Features

The goal of this study was to test whether labeling of one item’s feature would lead to the involuntary filtering of non-labeled features. We found asymmetric effects of labeling to the retention of the labeled and the non-labeled feature across different feature spaces. In Experiment 1, verbal labeling of orientation led to the filtering of continuous color. In contrast, color labeling did not result in substantial filtering of continuous orientation information (although a small cost was observed in recall error). Spatial frequency labeling in Experiment 2 led to the filtering of both continuous color and orientation memory. In contrast, spatial frequency information was not filtered with color labeling, and the effect of orientation labeling was ambiguous and almost negligible. In sum, color information is filtered with the labeling of any other visual feature (i.e., orientation and spatial frequency, and also shape as observed in Overkott and Souza, 2020), but color labeling does not result in the filtering of the non-labeled features. For the other visual features, the pattern of costs was more varied. Orientation labeling did not hamper spatial frequency memory, but spatial frequency labeling hampered orientation.

Why are some features more prone to filtering than others? The verbal labeling output in Experiment 2 showed that participants used far more labels (and hence had far more categorical knowledge) for colors than for the other features. Souza et al. (2020) showed that benefits of color labeling (and also shape labeling) on continuous imprecision depend on the number of labels used to describe the memoranda: with few terms (2 labels) memory got less precise, and with more terms (i.e., 7+ as used here), memory precision increased in comparison to the suppression condition. Interestingly, in the present study, orientation labeling showed a large benefit on orientation memory even if people used far fewer terms to describe this feature space (3 labels) compared to color. Therefore, the sensitivity of labeled and non-labeled features to labeling benefits and costs does not seem related to the number of terms used to describe the labeled feature.

One alternative explanation could be that the color feature is more costly to retain in memory in a continuous format. This possibility would be in line with the fact that, under suppression, participants could retain far less continuous information about colors than about orientations (see estimates in Table 3): across all experiments, mean estimates for continuous color memory under suppression were between 0.29 and 0.37, whereas continuous orientation memory estimates under suppression were between 0.41 and 0.48. Thus, color labeling may have helped to create a stable representation of the precise color hue that would be otherwise quickly lost. This would also explain the large susceptibility of color memory to costs due to the labeling of other features: if labeling of a feature requires heightened attentional processing of this feature, this would mean that less attention was used to process color, reducing its consolidation in memory.

The pattern of costs produced by labeling was not easily predicted based on just which feature was labeled vs. not-labeled: orientation memory was not much affected by color labeling, but continuous orientation memory was impaired when spatial frequency was labeled. This result suggests that depending on the exact pair of relevant features, directing attention to one of them may lead to the suppression of the other feature. People may be able to pay attention to color without losing the form, orientation or spatial frequency of the object, but an increase in the processing of these visual properties through labeling leads to the loss of color information.

10.5.2. Implications for Verbal Labeling Hypotheses

Across Experiments 1a, 1b and 2, we consistently observed that labeling increased the retention of continuous information about the labeled feature. This boost in continuous memory has been interpreted as reflecting a benefit caused by the activation of categorical knowledge in VLTM. To recapitulate, this hypothesis predicts that verbal labeling adds a second visual trace to the already formed visual trace in VWM. The label activates a category, which is then used as a reference for the creation of a precise visual memory trace. For example, if the category "blue" is activated it will be contrasted to the visual input and thereby information about the precise hue can be added in relation to the category. Without this reference, continuous information might be quickly lost.

The label as attentional-cue hypothesis assumes that the label guides attention to the labeled feature. According to Kelly and Heit (2017), this would only be beneficial when people were not already intentionally paying attention to the relevant feature of the memory object. To the extent that participants are fully aware of which feature is relevant for the memory test, this hypothesis predicts no labeling benefit to the labeled feature. However, if attention is withdrawn from the non-labeled feature, then this hypothesis would predict a cost to the non-labeled feature. Indeed, in one previous study in our lab, we observed that labeling of the object's shape produced a cost for the retention of color in VWM (Overkott & Souza, 2020), which we have interpreted as consistent with the attentional-cue hypothesis of labeling. Here the pattern of costs induced by labeling on the retention of some of the non-labeled features may be explained by differential attentional processing of these features.

10.5.3. Labeling vs. Attention to Features

The labeling costs observed here raise the question whether involuntary filtering is specific to labeling or it may be a by-product of increased attention to one feature dimension. Souza and Skóra (2017) assessed whether color labeling would yield similar performance to a condition in which participants made a preference judgment (like-dislike) on the presented color. Labeling produced better performance than preference rating which in turn produced comparable levels of performance to suppression. This result suggests that the labeling boost is more than just increased attention processing afforded by the dual-task requirement to label the color. However, it is unclear whether labeling costs could be simply related to increased attention to the labeled feature. Future studies may include control conditions that require attention processing of one feature dimension to assess whether similar costs will follow to the non-processed feature. For example, one could present a single stimulus varying on two features. Then, either one feature is labeled (as here), or the participants need to respond to a secondary task requiring attention to one feature but not the other. If performance costs are similar between the labeling and the attention condition, this would suggest that involuntary filtering is a by-product of attentional processing. In contrast, if only labeling produces costs this would indicate a specific contribution of verbal descriptions to the involuntary filtering of features.

10.5.4. Implications of Involuntary Filtering

Overall, the finding that labeling one feature sometimes resulted in the involuntary filtering of the non-labeled feature could imply that the labeling boost may involve the suppression of other information to more effectively gate the entrance of the labeled feature in VWM. This extends previous studies showing that item features can be filtered and hence independently retained in memory (Bays et al., 2011; Fougnie & Alvarez, 2011; Schneegans & Bays, 2017; Shin & Ma, 2017).

Furthermore, it points to trade-offs on how the limited VMW capacity (see, Cowan, 2010; Oberauer et al., 2016) is allocated to store different features, as not all item features were involuntarily filtered. The finding that certain item features were not filtered suggests that maintaining this feature required less attention as there was enough attention available to retain this feature in mind in addition to the labeled one. This suggests that orientation seems to require less attention than frequency and also color information, as it was the least likely to be filtered. In contrast, color seems to require more attention, is more costly to retain in memory and is filtered as soon as attention is directed to another feature through verbal labeling.

10.6. Conclusion

Verbal labels are assumed to activate categorical information in VLTM, thereby boosting the maintenance of high-fidelity information in VWM. Here, we found asymmetric labeling effects on the labeled and non-labeled features of an item: whereas the labeled feature always showed a benefit compared to suppression, some of the non-labeled features suffered when other features were labeled indicating that they were involuntarily filtered. Verbally describing our surroundings improves the detailed information we retain in memory about the described features a moment later – but, this can come at the expense of losing the information we did not describe.

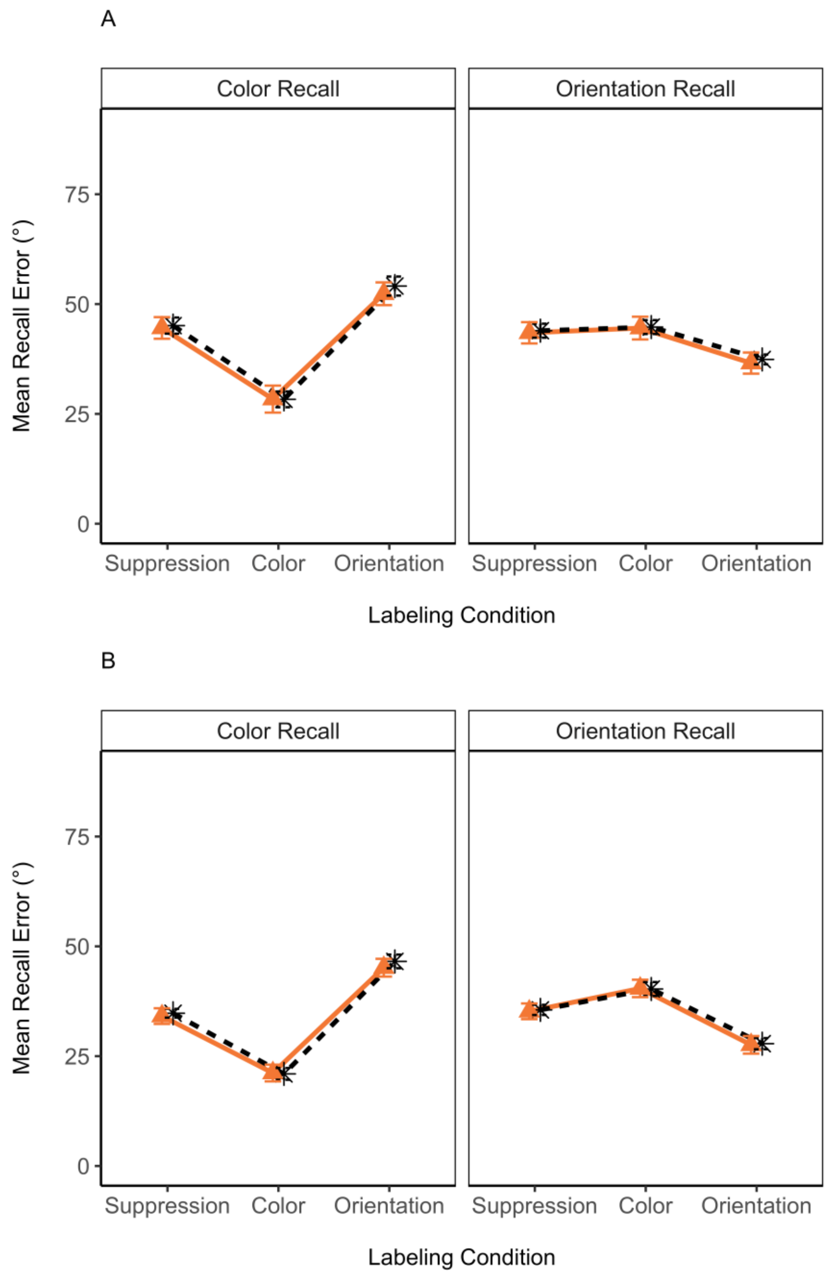
10.7. Appendix A

10.7.1. Model Fit

To assess how well the model captured the data, a posterior predictive check was performed by simulating data (predictions) based on the full model parameters for all experiments. Figure A1A shows that the predicted recall error seemed to be fairly in line with the data for Experiment 1a. Figure A1B shows that the modeling fit the data for Experiment 1b, but did this less well for orientation labeling under orientation recall. Figure A2 shows that the posterior estimates of the model in Experiment 2 also reproduced the actual data.

Figure A1

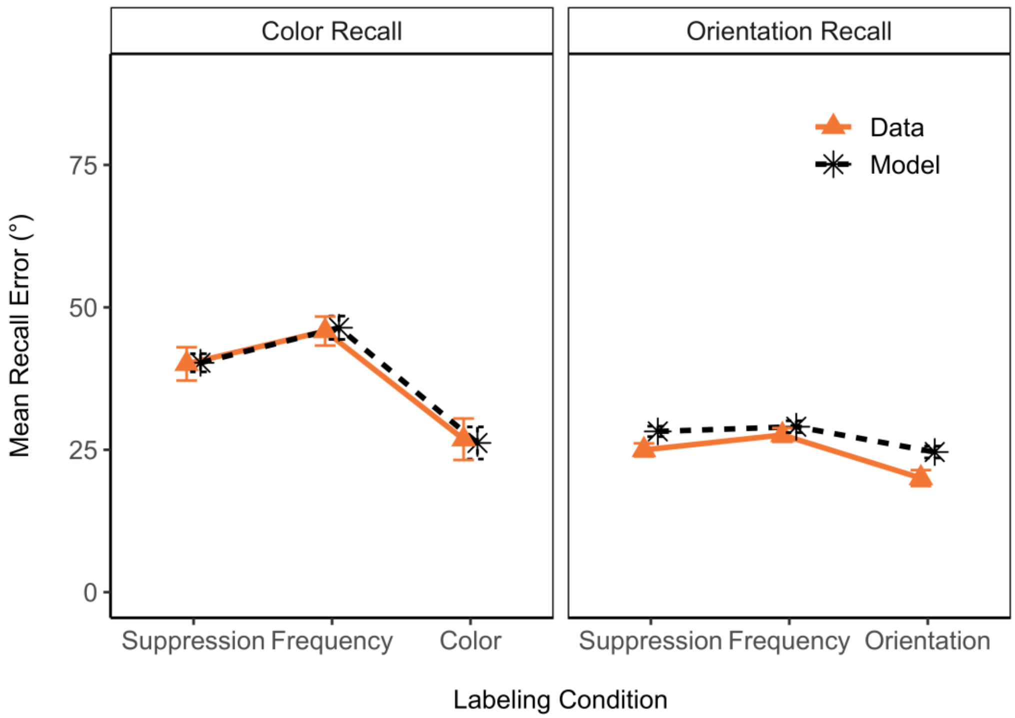
Recall Error Obtained for the Data of Experiment 1a (Panel A) and Experiment 1b (Panel B) and the Predicted, Simulated Data from the Posterior Estimates of the Mixture Model Fitted to this Data



Note. Error bars represent the 95% within-subjects confidence interval.

Figure A2

Recall Error Obtained for the Data of Experiment23 and the Predicted, Simulated Data from the Posterior Estimates of the Mixture Model Fitted to this Data



Note. Error bars represent the 95% within-subjects confidence interval. Color and Orientation indicate the type of recall test, whereas suppression and labeling indicate the labeling condition.

11. Curriculum Vitae

Clara Overkott

Nationality: Swiss | Date of birth: 17.12.1989 | |Mail: claraoverkott@gmail.com

Education

02/2017 – current	**PhD in Cognitive Psychology**, University or Zurich *Topics: (1) Consolidation in visual working and long-term memory and (2) Influence of verbalizations on visual working and long-term memory* *Supervisors: Dr. Alessandra S. Souza, Prof. Dr. Klaus Oberauer and Prof. Dr. Mark Nieuwenstein*
08/2013 – 01/2016	**Master of Science in Cognitive Psychology and Cognitive Neuroscience (major subject) and Biology**, University of Zurich *Thesis title: „The effects of reactivating long-term memory representations on working memory performance"*
01/2014	**Certificate "Swiss Association of Adult Education 1",** qualification in adult education, Switzerland
08/2009 – 07/2013	**Bachelor of Science in Psychology (major subject) and Biology** University of Zurich *Thesis title: "Verarbeitung von bekannten und unbekannten Gesichtern"*

Employment History

02/2017 – current	**PhD candidate and Junior Researcher** (60%), within the SNSF-funded project "The interaction of perception and language in memory: How do labels shape visual working memory?" to Dr. Alessandra S. Souza *Advisor: Dr. Alessandra S. Souza and Prof. Dr. Klaus Oberauer*
09/2019 – 02/2020	Visiting research assistant (100 %) at Cardiff University (Wales, UK), withing the SNSF-funded project "How do Verbal Labels Influence Children's Visual Working Memory?" to Clara Overkott *Advisor: Dr. Candice Morey*
09/2012 – 05/2017	Night auditory and secretary (20-50%), Sleep Laboratory Zurich Fluntern
02/2016 – 07/2016	**Research internship/ visiting student** (100%), School of Experimental Psychology – Cognitive Development, University of Bristol (UK), supported by the Swiss-European Mobility Programme (SEMP) *Advisor: Prof. Dr. Chris Jarrold* • Project leader of an experimental study • Programming of experiments, data analysis and interpretation
05/2014 – 01/2016	**Research** assistant (20-40%), Department of Cognitive Psychology, University of Zurich *Advisor: Dr. Claudia von Bastian and Dr. Carla De Simoni* • Study leader, part-time data collection project leader
08/2014 – 01/2015	Psychology internship (100% and 20%), Department of Child and Adolescent Psychiatry, University of Zurich
06/2012 – 03/2013	**Research internship** (15%), Department of Clinical Psychology and Psychotherapy, University of Zurich
01/2009 – 05/2009	Development cooperation internship (100%), with Nouvelle Planète and Maharogi Sewa Samiti in Nagpur, India (Rep. 2007, 2011)
10/2008 – 10/2014	Various student jobs (cashier, promotion assistant)

Institutional Responsibilities

10/2018 – current	Representative of the junior faculty for doctorates at the Department of Psychology – University of Zurich
06/2017 – current	Partially responsible for the general supervision and hiring of research assistants and tutors at the Department for Cognitive Psychology

Supervision of Students/ Junior Researcher

02/2017 – current	Supervision of various research assistants
10/2017 – 04/2018	Research intern: Marta Matyja (co-supervision)

Teaching Activities

01/2019 – 07/2019	Experimental Lab Course Students learn to conduct, process, analyze and interpret a psychological experiment
01/2018 – 07/2018	Experimental Lab Course

Active Memberships

Student Member of Psychonomic Society

Organization of Conferences

29/05/2018	Student Conference, students present their posters of their Experimental Lab Course, University of Zurich – Psychology Department

Prizes, Awards, Fellowships

02/2017 – 07/2017	Doc.Mobility Grant of the Swiss National Science Foundation (24'650 CHF)
05/2018	Poster prize at MaDoKo (Masterstudierenden und Doktorierenden Kongress), University of Zurich – Psychology Department (peer-reviewed)
02/2017 – 07/2017	Swiss-European Mobility Program grant for 6 months (2140 CHF)

Personal Skills

German	Mother tongue (C2)
English	Work language, fluent in writing and speaking (C2)
French	Fluent in writing and speaking (C2)
Hindi	Basic knowledge in writing and speaking (A1)

Programming	MATLAB, LATEX, JavaScript, HTML & CSS, LiveCode
Data Analysis	R, SPSS, JASP
Microsoft Office	Excel, Word, PowerPoint, Outlook
Adobe	InDesign, Illustrator, Photoshop
Methodologies	EEG, EKG, EMG, CPAP (continuous positive airway pressure)

Further Leadership and Project Management Experience

Regional manager of 7 scout divisions (840 people)
Expert (specialized course instructor) in camp-sport and trekking, Swiss Youth and Sports

Peer-reviewed Articles

Overkott, C., & Souza, A. S. (2020). Consolidation is not a ballistic process: Evidence that a distractor task disrupts ongoing consolidation in visual working memory. *Submitted Manuscript.*

Overkott, C., & Souza, A. S. (2020). The Fate of Labeled and Non-Labeled Visual Features in Working Memory. *Submitted Manuscript.*

Overkott, C., & Souza, A. S. (2020). Verbal Descriptions Improve Visual Working Memory, but Have Limited Impact on Episodic Visual Memory. *Submitted Manuscript.*

Souza, A. S., **Overkott, C.,** & Matyja, M. (2020). Categorical distinctiveness constrains the labeling benefit in visual working memory. *Submitted Manuscript.*

Oral Contributions to International Conferences

Overkott C., & Souza A. S. (2019, November). *Evidence Against All-Or-None Short-Term Consolidation: Consolidation Speeds up with Prioritization and Is Interrupted by a Distractor Task*. Poster presented at Psychonomic Socitey's 60st Annual Meeting, Montréal, Canada.

Souza A. S., **Overkott C.** , Matyja M. & Jang, J. (2019, November). *Improving Visual Working Memory with Verbal Labeling*. Psychonomic Socitey's 60st Annual Meeting, Montréal, Canada.

Overkott C. , Matyja M., & Souza A. S. (2019, September). *The Labeling Benefit in Visual Working Memory: How Specific do Verbal Labels Need to be?*. Poster presented at 21st Conference of the European Society for Cognitive Psychology (Escop), Tenerife, Spain.

Overkott, C., & Souza, A.S. (**2018**, November). *Is consolidation in visual working memory an all-or-none-process?* Poster presented at Psychonomic Socitey's 59th Annual Meeting , New Orleans, United States.

Souza, A.S., **Overkott, C. S. R.,** & Oberauer, K. (**2018**, November). *When does working memory for visuospatial arrays get better with repetitions? Unlocking the Hebb effect.* Psychonomic Socitey's 59th Annual Meeting, New Orleans, United States.

Overkott, C. S. R., & Souza, A.S. (**2018**, August). *Is consolidation an all-or-none process? Examining the interaction between encoding and recall order in visual working memory* Poster presented at the 9th European Working Memory Symposium, Pavia, Italy.

Souza, A.S., & **Overkott, C.** (**2018**, August). *From color terms to color memory: Does labeling in working memory boost episodic long-term memory?* Poster presented at the 9th European Working Memory Symposium, Pavia, Italy.

Overkott, C. S. R., & Souza, A. S. (**2017**, July). *Time to consolidate information in working memory and long-term memory.* SNSF International Exploratory Workshop on "The crossroads of attention in working memory: consolidation, refreshing and removal", Ovronnaz, Switzerland.

Overkott, C. S. R., & Souza, A. S. (**2017**, November). *Filtering of information in visual working memory: The effect of labelling object features.* Poster presented at Psychonomic Socitey's 58th Annual Meeting, Vancouver, Canada.

12. Acknowledgements

I would like to express my heartfelt gratitude to my main supervisor Dr. Alessandra S. Souza. Thank you for always supporting me with your advice, guidance and critical thinking. I would like to thank you for all the hard you have put in me so that I could follow into your steps and for always making me excited about my research.

I would further like to thank my official supervisor Prof. Dr. Klaus Oberauer for always being there when needed. Furthermore, I would like to thank Prof. Dr. Mark Nieuwenstein for being a member of the examining committee and reviewing this work.

I was very honored to have worked with Dr. Candice Morey at Cardiff University. Thank you for your advice, guidance and support. I would like to thank Teddy, Lizzie and Craig for making me feel at home in Cardiff. Thank you, Ali for being a social roommate and introducing me to the Cardiff University Triathlon Club – thank you all for showing me the beautiful landscape in Wales.

Many thanks go to my office colleagues Andra, Dave, Joscha and Samuel as well as Eda, Gidon, Julia and Lea for the coffee breaks. I would like to thank you all for the vivid discussions and your constant support. Furthermore, I thank all members of the Cognitive Psychology Unit for this pleasant time: Anke, Agnes, Christiane, Danielle, Emre, Hanna, Henrik, Hsuan-Yu, Jiri, Peter, Sizhu, Stefan, Suuad Tuule.

I would like to thank my parents Sandra and Luc for supporting me throughout my life. Special thanks go to my sister Dija and her family as well as my grandmother Stéphanie and her sister Primerose. Many thanks to my greatest friends Chiara and Isi, the ladies of the Chaostruppe in Chlosters and my friends of Pfadi Walter Tell.

Words cannot express how thankful I am for everything Roman has done for me.

13. References

Allen, R. J., & Ueno, T. (2018). Multiple high-reward items can be prioritized in working memory but with greater vulnerability to interference. *Attention, Perception, & Psychophysics*, *80*(7), 1731–1743. https://doi.org/10.3758/s13414-018-1543-6

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š., Birch, S., Birt, A. R., Bornstein, B. H., Bouwmeester, S., Brandimonte, M. A., & Brown, C. (2014). Registered replication report: Schooler and engstler-schooler (1990). *Perspectives on Psychological Science*, *9*(5), 556–578.

Atkinson, R. C., & Shiffrin, R. M. (1968). *Human memory: A proposed system and its control processes*.

Bae, G.-Y., & Flombaum, J. I. (2013). Two items remembered as precisely as one: How integral features can improve visual working memory. *Psychological Science*, *24*(10), 2038–2047.

Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*(4), 744.

Bayliss, D. M., Bogdanovs, J., & Jarrold, C. (2015). Consolidating working memory: Distinguishing the effects of consolidation, rehearsal and attentional refreshing in a working memory span task. *Journal of Memory and Language*, *81*, 34–50. https://doi.org/10.1016/j.jml.2014.12.004

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7–7. https://doi.org/10.1167/9.10.7

Bays, P. M., & Husain, M. (2008). Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science*, *321*(5890), 851–854. https://doi.org/10.1126/science.1158023

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, *49*(6), 1622–1631.

Becker, M. W., Miller, J. R., & Liu, T. (2013). A severe capacity limit in the consolidation of orientation information into visual short-term memory. *Attention, Perception, & Psychophysics*, *75*(3), 415–425. https://doi.org/10.3758/s13414-012-0410-0

Biderman, N., Luria, R., Teodorescu, A. R., Hajaj, R., & Goshen-Gottstein, Y. (2019). Working Memory Has Better Fidelity Than Long-Term Memory: The Fidelity Constraint Is Not a General Property of Memory After All. *Psychological Science*, *30*(2), 223–237. https://doi.org/10.1177/0956797618813538

Blanco, N., & Gureckis, T. (2013). Does category labeling lead to forgetting? *Cognitive Processing*, *14*(1), 73–79. https://doi.org/10.1007/s10339-012-0530-4

Bocincova, A., & Johnson, J. S. (2019). The time course of encoding and maintenance of task-relevant versus irrelevant object features in working memory. *Cortex*, *111*, 196–209. https://doi.org/10.1016/j.cortex.2018.10.013

Bocincova, A., van Lamsweerde, A. E., & Johnson, J. S. (2017). The role of top-down suppression in mitigating the disruptive effects of task-irrelevant feature changes in visual working memory. *Memory & Cognition*, *45*(8), 1411–1422.

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *Journal of Neuroscience*, *35*(25), 9329–9335.

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review*, *114*(1), 38.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical Encoding in Visual Working Memory: Ensemble Statistics Bias Memory for Individual Items. *Psychological Science*, *22*(3), 384–392. https://doi.org/10.1177/0956797610397956

Brady, T. F., & Alvarez, G. A. (2015a). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*, *15*(15), 1–24.

Brady, T. F., & Alvarez, G. A. (2015b). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 921–929. https://doi.org/10.1037/xlm0000075

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*(4), 487–502. https://doi.org/10.1037/a0016797

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 4. https://doi.org/10.1167/11.5.4

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual Long-Term Memory Has a Massive Storage Capacity for Object Details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14325–14329.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013). Real-world objects are not represented as bound units: Independent forgetting of different object details from visual memory. *Journal of Experimental Psychology: General*, *142*(3), 791–808. https://doi.org/10.1037/a0029649

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, *24*(6), 981–990.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. M. (1992). Verbal recoding of visual stimuli impairs mentalimagetransformations. *Memory & Cognition*, *20*(4), 449–455. https://doi.org/10.3758/BF03210929

Brandimonte, M. A., Schooler, J. W., & Gabbino, P. (1997). Attenuating verbal overshadowing through color retrieval cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 915–931. https://doi.org/10.1037/0278-7393.23.4.915

Brown, C., Brandimonte, M. A., Wickham, L. H., Bosco, A., & Schooler, J. W. (2014). When do words hurt? A multiprocess view of the effects of verbalization on visual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1244.

Chekaf, M., Cowan, N., & Mathy, F. (2016). Chunk formation in immediate memory and how it relates to data compression. *Cognition*, *155*, 96–107.

Chen, H., & Wyble, B. (2015). The location but not the attributes of visual cues are automatically encoded into working memory. *Vision Research*, *107*, 76–85.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology. Human Perception and Performance*, *21*(1), 109–127.

Couture, M., & Tremblay, S. (2006). Exploring the characteristics of the visuospatial hebb repetition effect. *Memory & Cognition*, *34*(8), 1720–1729. https://doi.org/10.3758/BF03195933

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87–114; discussion 114-185.

Cowan, Nelson. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163.

Cowan, Nelson. (2008). Chapter 20 What are the differences between long-term, short-term, and working memory? In W. S. Sossin, J.-C. Lacaille, V. F. Castellucci, &

S. Belleville (Eds.), *Progress in Brain Research* (Vol. 169, pp. 323–338). Elsevier. https://doi.org/10.1016/S0079-6123(07)00020-9

Cowan, Nelson. (2010). The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Cowan, Nelson. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Cowan, Nelson, Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 731.

De Schrijver, S., & Barrouillet, P. (2017). Consolidation and restoration of memory traces in working memory. *Psychonomic Bulletin & Review*, 1–7.

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review*, *22*(1), 170–178.

Dux, P. E., & Marois, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception, & Psychophysics*, *71*(8), 1683–1700. https://doi.org/10.3758/APP.71.8.1683

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93–100. https://doi.org/10.1016/j.cognition.2015.06.008

Fischer, R., & Plessow, F. (2015). Efficient multitasking: Parallel versus serial processing of multiple tasks. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01366

Foerster, R. M., & Schneider, W. X. (2018). Involuntary top-down control by search-irrelevant features: Visual working memory biases attention in an object-based manner. *Cognition*, *172*, 37–45. https://doi.org/10.1016/j.cognition.2017.12.002

Forsberg, A., Johnson, W., & Logie, R. H. (2019). Aging and feature-binding in visual working memory: The role of verbal rehearsal. *Psychology and Aging*, *34*(7), 933–953. https://doi.org/10.1037/pag0000391

Forsberg, A., Johnson, W., & Logie, R. H. (2020). Cognitive aging and verbal labeling in continuous visual memory. *Memory & Cognition*.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision*, *11*(12), 3–3. https://doi.org/10.1167/11.12.3

Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, *10*(12), 27–27. https://doi.org/10.1167/10.12.27

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*(2), 177–182. https://doi.org/10.1016/j.conb.2010.03.005

Gajewski, D. A., & Brockmole, J. R. (2006). Feature bindings endure without attention: Evidence from an explicit recall task. *Psychonomic Bulletin & Review*, *13*(4), 581–587. https://doi.org/10.3758/BF03193966

Gao, Z., Yu, S., Zhu, C., Shui, R., Weng, X., Li, P., & Shen, M. (2016). Object-based Encoding in Visual Working Memory: Evidence from Memory-driven Attentional Capture. *Scientific Reports*, *6*(1), 22822. https://doi.org/10.1038/srep22822

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016.

Hardman, K. O. (2016). *CatContModel: Categorical and Continuous Working Memory Models for Delayed Estimation Tasks* (0.7.5) [Computer Software]. https://github.com/hardmanko/CatContModel/releases/tag/v0.7.5

Hardman, K. O., & Cowan, N. (2015). Remembering Complex Objects in Visual Working Memory: Do Capacity Limits Restrict Objects or Features? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *41*(2), 325–347. https://doi.org/10.1037/xlm0000031

Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 30–54. https://doi.org/10.1037/xhp0000290

Hu, Y., Hitch, G. J., Baddeley, A. D., Zhang, M., & Allen, R. J. (2014). Executive and perceptual attention play different roles in visual working memory: Evidence from suffix and strategy effects. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1665–1678. https://doi.org/10.1037/a0037163

Huang, L., & Awh, E. (2018). Chunking in working memory via content-free labels. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-017-18157-5

Huebner, G. M., & Gegenfurtner, K. R. (2011). The efficiency of encoding: Limits of information transfer into memory. *Attention, Perception, & Psychophysics*, *73*(5), 1503–1521.

Jeffreys, H. (1961). *Theory of probability.,(Oxford University Press: Oxford, UK)*.

Johnson, A. J., Dygacz, A., & Miles, C. (2017). Hebb repetition effects for non-verbal visual sequences: Determinants of sequence acquisition. *Memory*, *25*(9), 1279–1293.

Jolicœur, P., & Dell'Acqua, R. (1998). The Demonstration of Short-Term Consolidation. *Cognitive Psychology*, *36*(2), 138–202. https://doi.org/10.1006/cogp.1998.0684

Kandemir, G., Akyürek, E. G., & Nieuwenstein, M. R. (2017). Retro-Active Emotion: Do Negative Emotional Stimuli Disrupt Consolidation in Working Memory? *PLOS ONE*, *12*(1), e0169927. https://doi.org/10.1371/journal.pone.0169927

Kelly, L. J., & Heit, E. (2017). Recognition memory for hue: Prototypical bias and the role of labeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 955.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Scene Memory Is More Detailed Than You Think: The Role of Categories in Visual Long-Term Memory. *Psychological Science*, *21*(11), 1551–1556.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*(3), 558–578. https://doi.org/10.1037/a0019165

Lafond, D., Tremblay, S., & Parmentier, F. (2010). The ubiquitous nature of the Hebb repetition effect: Error learning mistaken for the absence of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 515.

Lehle, C., & Hübner, R. (2009). Strategic capacity sharing between two tasks: Evidence from tasks with the same and with different task sets. *Psychological Research PRPF*, *73*(5), 707.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Lupyan, G. (2008). From chair to" chair": A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, *137*(2), 348.

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*, 54.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170.

Luria, R., & Vogel, E. K. (2011). Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia*, *49*(6), 1632–1639. https://doi.org/10.1016/j.neuropsychologia.2010.11.031

MacLeod, C. M. (2010). When learning met memory. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *64*(4), 227–240. https://doi.org/10.1037/a0021699

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671.

Makovski, T. (2016). Does proactive interference play a significant role in visual working memory tasks? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1664–1672. https://doi.org/10.1037/xlm0000262

Maniglia, M. R., & Souza, A. S. (2020). Age Differences in the Efficiency of Filtering and Ignoring Distraction in Working Memory. *Manuscript Submitted for Publication*.

Marshall, L., & Bays, P. M. (2013). Obligatory encoding of task-irrelevant features depletes working memory resources. *Journal of Vision*, *13*(2), 21–21.

Massaro, D. W. (1975). Backward recognition masking. *The Journal of the Acoustical Society of America*, *58*(5), 1059–1065. https://doi.org/10.1121/1.380765

Miller, J. R., Becker, M. W., & Liu, T. (2014). The bandwidth of consolidation into visual short-term memory depends on the visual feature. *Visual Cognition*, *22*(7), 920–947. https://doi.org/10.1080/13506285.2014.936923

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes Factors for Common Designs.* (R package version 0.9.12-2) [Computer software].

Nassar, M. R., Helmers, J. C., & Frank, M. J. (2018). Chunking as a rational strategy for lossy data compression in visual working memory. *Psychological Review*, *125*(4), 486.

Nieuwenstein, M., & Wyble, B. (2014). Beyond a mask and against the bottleneck: Retroactive dual-task interference during working memory consolidation of a masked visual target. *Journal of Experimental Psychology: General*, *143*(3), 1409.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411–421. https://doi.org/10.1037/0278-7393.28.3.411

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation*, *51*, 45–100.

Oberauer, K. (2019). Working Memory and Attention – A Conceptual Analysis and Review. *Journal of Cognition*, *2*(1), 36. https://doi.org/10.5334/joc.58

Oberauer, K., Awh, E., & Sutterer, D. W. (2017). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 1–22. https://doi.org/10.1037/xlm0000302

Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Memory & Cognition*, *41*(8), 1212–1227. https://doi.org/10.3758/s13421-013-0333-6

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799. https://doi.org/10.1037/bul0000046

Oberauer, K., & Hein, L. (2012). Attention to Information in Working Memory. *Current Directions in Psychological Science*, *21*(3), 164–169. https://doi.org/10.1177/0963721412444727

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*(1), 21–59. https://doi.org/10.1037/rev0000044

Oberauer, K., & Meyer, N. (2009). The contributions of encoding, retention, and recall to the Hebb effect. *Memory*, *17*(7), 774–781.

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Perception & Psychophysics*, *64*(7), 1055–1067. https://doi.org/10.3758/BF03194756

Overkott, C., & Souza, A. S. (2020). Verbal Descriptions Improve Visual Working Memory, but Have Limited Impact on Episodic Visual Memory. *Submitted Manuscript*.

Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1543–1547. https://doi.org/10.1037/a0020604

Palmer, J., Boston, B., & Moore, C. M. (2015). Limited capacity for memory tasks with multiple features within a single object. *Attention, Perception, & Psychophysics*, *77*(5), 1488–1499. https://doi.org/10.3758/s13414-015-0909-2

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Persaud, K., & Hemmer, P. (2016). The dynamics of fidelity over the time course of long-term memory. *Cognitive Psychology*, *88*, 1–21. https://doi.org/10.1016/j.cogpsych.2016.05.003

Pratte, M. S., Park, Y. E., Rademaker, R. L., & Tong, F. (2017). Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 6–17. https://doi.org/10.1037/xhp0000302

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998a). Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 261.

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998b). Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 261–282. https://doi.org/10.1037/0096-1523.24.1.261

Quak, M., Langford, Z. D., London, R. E., & Talsma, D. (2018). Contralateral delay activity does not reflect behavioral feature load in visual working memory. *Biological Psychology*, *137*, 107–115.

R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Richler, J. J., Gauthier, I., & Palmeri, T. J. (2011). Automaticity of basic-level categorization accounts for labeling effects in visual recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1579–1587. https://doi.org/10.1037/a0024347

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2013). How does using object names influence visual recognition memory? *Journal of Memory and Language*, *68*(1), 10–25.

Ricker, T. J. (2015). The Role of Short-term Consolidation in Memory Persistence. AIMS Neuroscience, 2015, 2 (4): 259-279. doi: 10.3934. *Neuroscience*, *259*. http://www.aimspress.com/article/10.3934/Neuroscience.2015.4.259

Ricker, T. J., & Cowan, N. (2014). Differences between presentation methods in working memory procedures: A matter of working memory consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 417.

Ricker, T. J., & Hardman, K. O. (2017). The nature of short-term consolidation in visual working memory. *Journal of Experimental Psychology. General*, *146*(11), 1551–1573. https://doi.org/10.1037/xge0000346

Ricker, T. J., Nieuwenstein, M. R., Bayliss, D. M., & Barrouillet, P. (2018). Working memory consolidation: Insights from studies on attention and working memory. *Annals of the New York Academy of Sciences*.

Ricker, T. J., & Sandry, J. (2018). The relationship between masking and short-term consolidation during recall from visual working memory. *Annals of the New York Academy of Sciences*.

Rideaux, R., Baker, E., & Edwards, M. (2018). Parallel consolidation into visual working memory results in reduced precision representations. *Vision Research*, *149*, 24–29. https://doi.org/10.1016/j.visres.2018.06.005

Rock, I., Linnett, C. M., Grant, P., & Mack, A. (1992). Perception without attention: Results of a new method. *Cognitive Psychology*, *24*(4), 502–534.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J. D. (2006). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, *1*(3), 181–210.

Roediger, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition*, *1*(4), 242–248. https://doi.org/10.1016/j.jarmac.2012.09.002

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374.

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E.-J. (2017). Bayesian analysis of factorial designs. *Psychological Methods*, *22*(2), 304.

Schneegans, S., & Bays, P. M. (2017). Neural Architecture for Feature Binding in Visual Working Memory. *Journal of Neuroscience*, *37*(14), 3913–3925. https://doi.org/10.1523/JNEUROSCI.3493-16.2017

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*(1), 36–71.

Sense, F., Morey, C. C., Prince, M., Heathcote, A., & Morey, R. D. (2017). Opportunity for verbalization does not improve visual change detection performance: A state-trace analysis. *Behavior Research Methods*, *49*(3), 853–862. https://doi.org/10.3758/s13428-016-0741-1

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-Specific Delay Activity in Human Primary Visual Cortex. *Psychological Science*, *20*(2), 207–214. https://doi.org/10.1111/j.1467-9280.2009.02276.x

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*(8), 291–296. https://doi.org/10.1016/S1364-6613(97)01094-2

Shin, H., & Ma, W. J. (2016). Crowdsourced single-trial probes of visual working memory for irrelevant features. *Journal of Vision*, *16*(5), 10–10. https://doi.org/10.1167/16.5.10

Shin, H., & Ma, W. J. (2017). Visual short-term memory for oriented, colored objects. *Journal of Vision*, *17*(9), 12–12. https://doi.org/10.1167/17.9.12

Souza, A. S., Overkott, C., & Matyja, M. (2020). Categorical distinctiveness constrains the labeling benefit in visual working memory. *Submitted Manuscript*.

Souza, A. S., Rerko, L., & Oberauer, K. (2016). Getting more from visual working memory: Retro-cues enhance retrieval and protect from visual interference.

Journal of Experimental Psychology: Human Perception and Performance, *42*(6), 890.

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, *166*, 277–297. https://doi.org/10.1016/j.cognition.2017.05.038

Souza, A. S., & Vergauwe, E. (2018). Unravelling the intersections between consolidation, refreshing, and removal. *Annals of the New York Academy of Sciences*, *1424*(1), 5–7. https://doi.org/10.1111/nyas.13943

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1.

Stevanovski, B., & Jolicœur, P. (2007). Visual short-term memory: Central capacity limitations in short-term consolidation. *Visual Cognition*, *15*(5), 532–563.

Sutterer, D. W., & Awh, E. (2016). Retrieval practice enhances the accessibility but not the quality of memory. *Psychonomic Bulletin & Review*, *23*(3), 831–841. https://doi.org/10.3758/s13423-015-0937-x

Swan, G., Collins, J., & Wyble, B. (2016). Memory for a single object has differently variable precisions for relevant and irrelevant features. *Journal of Vision*, *16*(3), 32–32. https://doi.org/10.1167/16.3.32

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics, Always Learning*. Boston, MA: Pearson.

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology.*, *45*, 37–55. https://doi.org/10.1037/xlm0000578

Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 3.

Tombu, M. N., Asplund, C. L., Dux, P. E., Godwin, D., Martin, J. W., & Marois, R. (2011). A unified attentional bottleneck in the human brain. *Proceedings of the National Academy of Sciences*, *108*(33), 13426–13431.

Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition*, *34*(8), 1704–1719. https://doi.org/10.3758/BF03195932

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, *80*(1), 1–52. https://doi.org/10.1037/h0033872

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780–8785. https://doi.org/10.1073/pnas.1117465109

Vergauwe, E., Ricker, T. J., Langerock, N., & Cowan, N. (2019). What do people typically do between list items? The nature of attention-based mnemonic activities depends on task context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(5), 779–794. https://doi.org/10.1037/xlm0000625

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436–1451. https://doi.org/10.1037/0096-1523.32.6.1436

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspectives on Psychological Science*, *6*(3), 291–298.

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*(6), 1057–1064.

Wilken, P., & Ma, W. J. (2004a). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11–11.

Wilken, P., & Ma, W. J. (2004b). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11. https://doi.org/10.1167/4.12.11

Woodman, G. F., & Vogel, E. K. (2005). Fractionating working memory: Consolidation and maintenance are independent processes. *Psychological Science*, *16*(2), 106–113.

Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review*, *15*(1), 223–229. https://doi.org/10.3758/PBR.15.1.223

Wyble, B., Bowman, H., & Nieuwenstein, M. (2009). The attentional blink provides episodic distinctiveness: Sparing at a cost. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 787.

Wyble, B., Bowman, H., & Nieuwenstein, M. (2015). On the interplay between working memory consolidation and attentional selection in controlling conscious access: Parallel processing at a cost—a comment on 'The interplay of attention and consciousness in visual search, attentional blink and working memory consolidation.' *Phil. Trans. R. Soc. B*, *370*(1661), 20140197.

Wyble, B., Potter, M. C., Bowman, H., & Nieuwenstein, M. (2011). Attentional episodes in visual perception. *Journal of Experimental Psychology: General*, *140*(3), 488.

Yu, Q., & Shim, W. M. (2017). Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working memory. *NeuroImage*, *157*, 97–107. https://doi.org/10.1016/j.neuroimage.2017.05.055

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Zylberberg, A., Dehaene, S., Roelfsema, P. R., & Sigman, M. (2011). The human Turing machine: A neural framework for mental programs. *Trends in Cognitive Sciences*, *15*(7), 293–300.